

---

---

# Prognostic Value of $^{18}\text{F}$ -FDG PET Radiomics Features at Baseline in PET-Guided Consolidation Strategy in Diffuse Large B-Cell Lymphoma: A Machine-Learning Analysis from the GAINED Study

Thomas Carlier<sup>1,2</sup>, Gauthier Frécon<sup>1,2</sup>, Diana Mateus<sup>3</sup>, Mira Rizkallah<sup>3</sup>, Françoise Kraeber-Bodéré<sup>1,2</sup>, Salim Kanoun<sup>4</sup>, Paul Blanc-Durand<sup>5</sup>, Emmanuel Itti<sup>5</sup>, Steven Le Gouill<sup>6</sup>, René-Olivier Casasnovas<sup>7</sup>, Caroline Bodet-Milin<sup>1,2</sup>, and Clément Bailly<sup>1,2</sup>

<sup>1</sup>Nantes Université, INSERM, CNRS, CRCF<sup>2</sup>NA, Université d'Angers, Nantes, France; <sup>2</sup>Nuclear Medicine Department, University Hospital, Nantes, France; <sup>3</sup>Laboratoire des Sciences Numériques de Nantes, Ecole Centrale de Nantes, CNRS UMR 6004, Nantes, France; <sup>4</sup>Nuclear Medicine, Georges-François Leclerc Center, Dijon, France; <sup>5</sup>Nuclear Medicine, CHU Henri Mondor, Paris-Est University, Créteil, France; <sup>6</sup>Haematology Department, University Hospital, Nantes, France; and <sup>7</sup>Hematology, CHU Dijon Bourgogne, Dijon, France

---

The results of the GA in Newly Diagnosed Diffuse Large B-Cell Lymphoma (GAINED) study demonstrated the success of an  $^{18}\text{F}$ -FDG PET-driven approach to allow early identification—for intensification therapy—of diffuse large B-cell lymphoma patients with a high risk of relapse. Besides, some works have reported the prognostic value of baseline PET radiomics features (RFs). This work investigated the added value of such biomarkers on survival of patients involved in the GAINED protocol. **Methods:** Conventional PET features and RFs were computed from  $^{18}\text{F}$ -FDG PET at baseline and extracted using different volume definitions (patient level, largest lesion, and hottest lesion). Clinical features and the consolidation treatment information were also considered in the model. Two machine-learning pipelines were trained with 80% of patients and tested on the remaining 20%. The training was repeated 100 times to highlight the test set variability. For the 2-y progression-free survival (PFS) outcome, the pipeline included a data augmentation and an elastic net logistic regression model. Results for different feature groups were compared using the mean area under the curve (AUC). For the survival outcome, the pipeline included a Cox univariate model to select the features. Then, the model included a split between high- and low-risk patients using the median of a regression score based on the coefficients of a penalized Cox multivariate approach. The log-rank test *P* values over the 100 loops were compared with a Wilcoxon signed-ranked test. **Results:** In total, 545 patients were included for the 2-y PFS classification and 561 for survival analysis. Clinical features alone, consolidation features alone, conventional PET features, and RFs extracted at patient level achieved an AUC of, respectively,  $0.65 \pm 0.07$ ,  $0.64 \pm 0.06$ ,  $0.60 \pm 0.07$ , and  $0.62 \pm 0.07$  ( $0.62 \pm 0.07$  for the largest lesion and  $0.54 \pm 0.07$  for the hottest). Combining clinical features with the consolidation features led to the best AUC ( $0.72 \pm 0.06$ ). Adding conventional PET features or RFs did not improve the results. For survival, the log-rank *P* values of the model involving clinical and consolidation features together were significantly smaller than all combined-feature groups (*P* < 0.007). **Conclusion:** The results showed that a concatenation of multimodal features coupled with a simple machine-learning model

does not seem to improve the results in terms of 2-y PFS classification and PFS prediction for patient treated according to the GAINED protocol.

**Key Words:** DLBCL;  $^{18}\text{F}$ -FDG PET; GAINED study; radiomics

**J Nucl Med 2024; 65:156–162**

DOI: 10.2967/jnumed.123.265872

---

**T**he use of PET with  $^{18}\text{F}$ -FDG is now considered the standard imaging procedure not only for staging but also for therapeutic response assessment of patients with diffuse large B-cell lymphoma (DLBCL) (1). Yet, the number of patients who do not respond after standard front-line therapy—that is, a combination of anti-CD20 monoclonal antibody with cyclophosphamide, doxorubicin, vincristine, and prednisone chemotherapy—is still significant (30%–40%) (2) and advocates for the use of new biomarkers to determine patients with a high risk of relapse. In that respect, approaches based on interim PET to identify these patients and to adapt consolidation strategies accordingly have been explored for over 10 y (3). Several studies also evaluated the potential benefit of using imaging biomarkers extracted from PET volumes (4) in combination (or not) with other clinical or demographic features at baseline. Most of the studies focused primarily on assessment of the SUV extracted from the most intense lesion ( $\text{SUV}_{\text{max}}$ ), the total metabolic tumor volume (TMTV), and dissemination. Although the role of  $\text{SUV}_{\text{max}}$  at baseline to predict 2-y progression-free survival (PFS) or overall survival was shown to be limited (4), TMTV (5) and dissemination (6) appeared to be promising at baseline (7).

Another area related to the high-throughput extraction of quantitative data from medical images has gained importance by considering the measurement of the spatial heterogeneity within a tumor (hereafter called radiomics) (8). Evaluation of the predictive values of these other potential image-based biomarkers in DLBCL patients has recently been considered (9–17). However, the usefulness of complex radiomics in this context remains controversial, as some studies reported a potential added value (9–11,14–17) whereas others concluded that there is a lack of, or only a moderate,

---

Received Apr. 14, 2023; revision accepted Oct. 17, 2023.  
For correspondence or reprints, contact Clément Bailly (clement.bailly@chu-nantes.fr).  
Published online Nov. 9, 2023.  
COPYRIGHT © 2024 by the Society of Nuclear Medicine and Molecular Imaging.

predictive value for 2-y PFS when these biomarkers are considered alone or combined (12,13). These contradictory results may be explained partly by the prospective or retrospective nature of each study, the population size and type, the presence (or not) of a test dataset, the monocentric or multicentric characteristic of the study, the heterogeneity in immunochemotherapy regimen and patient management, and the features considered in the model.

The GA in Newly Diagnosed Diffuse Large B-Cell Lymphoma (GAINED) randomized phase III trial was the first study demonstrating the success of a PET-driven strategy. Interim PET interpretation criteria were simple and based on  $SUV_{max}$  reduction ( $\Delta SUV_{max}$ ) (18). The purpose of our work was to explore the prognostic added value on 2-y PFS and survival (using PFS as the endpoint) of a combination of clinical, PET-based, and radiomics features (RFs) extracted at baseline in DLBCL patients included in this latter prospective multicenter cohort in which patient management is driven by PET follow-up. A secondary objective was to identify whether baseline PET characteristics could predict patients' response after 4 cycles of chemotherapy and consequently the treatment arm to which they would be assigned. Finally, a third was to assess whether a specific volume of interest (VOI) was more relevant for radiomics computation when the 2-y PFS classification and treatment prediction are considered.

## MATERIALS AND METHODS

### Patients and PET/CT Imaging

The GAINED phase 3 trial (NCT 01659099) was conducted in 99 centers and enrolled newly diagnosed untreated DLBCL patients between 18 and 60 y old. The main exclusion and inclusion criteria have already been reported (18). The patients received cyclophosphamide, doxorubicin, vincristine, and prednisone or ACVBP (doxorubicin, cyclophosphamide, vindesine, bleomycin, and prednisone) on a 14-d schedule plus obinutuzumab (GA-101) or rituximab. Interim PET was performed at diagnosis and after 2 and 4 cycles of chemotherapy (respectively, PET2 and PET4) and analyzed according to  $\Delta SUV_{max}$ .

Patients with an early good response (defined as negative PET2 and PET4) continued the planned immunochemotherapy; slow responders (positive PET2 and negative PET4) received intensification therapy with 2 courses of high-dose methotrexate followed by autologous stem cell transplantation. Patients with positive PET after 4 induction cycles received salvage therapy. No difference in terms of 2-y PFS was reported between the 2 induction arms (GA-101 vs. rituximab). Early good responders and slow responders had similar 2-y PFS and overall survival.  $^{18}F$ -FDG PET scans were performed according to the local procedure of each center. All images were reconstructed in SUV normalized for body weight.

### Segmentation, Features Processing, and Selection

A gross VOI was manually delineated for each tumor, removing nontumor adjacent regions when needed. The final segmentation for each tumor was extracted using a fixed SUV threshold of 4.0 (19). Several PET-based imaging biomarkers were extracted from each reconstructed volume using PyRadiomics version 3.0 (20). Four conventional PET features were extracted: TMTV, whole-body total lesion glycolysis,  $SUV_{max}$ , and dissemination (maximum distance between 2 lesions normalized by body surface area). Among the RFs available for computation, a subset of 39 features was preselected (Supplemental Fig. 1; supplemental materials are available at <http://jnm.snmjournals.org>), choosing those that presented the best properties of repeatability (21–23). Before RF computation, PET images were resampled to the same voxel size ( $2 \times 2 \times 2$  mm) using a bicubic spline interpolation. Data were subsequently normalized using 2 approaches: a linear equalization using 64 bins and a fixed bin width of SUV 0.3. A subsequent preprocessing step was used to enhance imaging characteristics: a wavelet transform using coiflet-1 filters to decompose the original image into 8 decomposition frequencies and an edge-enhancement Laplacian of gaussian filter (with 2- and 6-mm  $\sigma$ ). This led a total of 647 RFs, which were computed from 3 different VOIs: the VOI with the most intense  $SUV_{max}$  (named hottest), the VOI that corresponded to the largest volume (named largest), and TMTV.

The RF selection step was divided into 3 parts. First, the reliability of textural RFs (excluding first-order and shape features) was derived

**TABLE 1**  
Description of Different Models

Model no.	Model name	Associated features
1	aalPI	aalPI
2	Consolidation	Chemotherapy regimen, autologous cell transplantation, salvage therapy
3	Clinical	Age, Ann Arbor stage, ECOG status, number of extranodal sites, LDH
4	Conventional PET	TMTV, total lesion glycolysis, $SUV_{max}$ , sDmax
5	Radiomics PET	Selected RFs after selection step using largest lesion
6	Consolidation + clinical (models 2 + 3)	All features from clinical and consolidation
7	Consolidation + clinical + conventional PET (models 2 + 3 + 4)	All features from clinical, conventional (PET), and consolidation
8	Consolidation + clinical + radiomics PET (models 2 + 3 + 5)	All features from clinical, radiomics (PET), and consolidation
9	Consolidation + clinical + conventional PET + radiomics PET (models 2 + 3 + 4 + 5)	All features from clinical and conventional (PET), radiomics (PET), and consolidation

aalPI = age-adjusted international prognostic index; ECOG = Eastern Cooperative Oncology Group; LDH = lactate dehydrogenase; sDmax = maximum distance between 2 lesions normalized by body surface area.

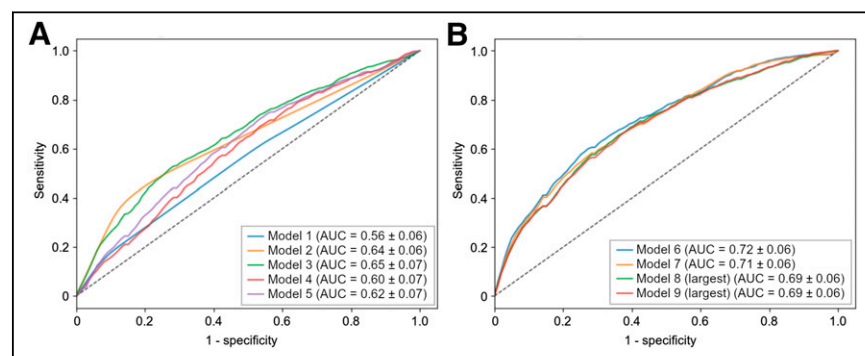
**TABLE 2**  
Patient Characteristics

Characteristic	2-y PFS classification (n = 545)	Survival (n = 561)
Events	86 (100%)	107 (100%)
Age (y)	48 (18–60)	48 (18–60)
Sex, male	305 (56%)	314 (56%)
aalPI		
0–1	230 (42%)	242 (43%)
2–3	315 (58%)	319 (57%)
Ann Arbor stage		
I–II	102 (19%)	106 (19%)
III–IV	443 (81%)	455 (81%)
Extranodal involvement		
<2	259 (48%)	269 (48%)
≥2	286 (52%)	292 (52%)
Performance status		
0–1	470 (86%)	485 (86%)
>1	75 (14%)	76 (14%)
LDH		
≤Normal	147 (27%)	155 (28%)
>Normal	398 (73%)	406 (72%)
Treatment arm		
GA-101	278 (51%)	286 (51%)
Rituximab	267 (49%)	275 (49%)
Induction treatment		
GA-101-CHOP	142 (26%)	145 (26%)
Rituximab-CHOP	144 (26%)	148 (26%)
GA-101-ACVBP	136 (25%)	141 (25%)
Rituximab-ACVBP	123 (23%)	127 (23%)

aalPI = age-adjusted international prognostic index; CHOP = cyclophosphamide, doxorubicin, vincristine, and prednisone.

Qualitative data are number and percentage; continuous data are mean and range.

following the methodology proposed by Pfähler et al. (24). Briefly, for each patient and each RF, the signal within the ROI was randomly shuffled 50 times and RF subsequently computed. The RF was considered unreliable if its original value lay within the 95% CI computed



**FIGURE 1.** Mean ROC curves for models 1–5 (A) and 6–9 (B) for 2-y PFS classification.

using the random shuffling. Finally, the RF was kept if its original value was outside the proposed CI for 90% of the patients. Second, all RFs that correlated strongly (Spearman correlation higher than 0.9) with 1 of the 4 conventional PET features were removed. Lastly, a hierarchic agglomerative clustering was applied to manage multicollinearity between features (supplemental methods).

All features were then standardized, setting the mean to zero and the SD to 1 to make the feature space homogeneous.

### Machine-Learning Models

The population was split into training and testing datasets with a ratio of 80% to 20%, stratified according to the 2-y PFS outcome for the 2-y PFS classification objective and according to the censorship and number of events per quartile for the survival analysis. Data augmentation was used on the training dataset to increase the minority class cardinal to the cardinal of the majority class. Both the synthetic minority oversampling technique (25) and the adaptive synthetic sampling approach for imbalanced learning (26) were considered one of the hyperparameters of the model.

For the 2-y PFS classification objective and the PET4 prediction, a logistic regression (LR) including elastic net regularization was considered. A 4-fold cross validation for the grid search was used on the training dataset (parameters detailed in Supplemental Table 1), with an objective of maximizing the area under the receiver operator characteristic (ROC) curve (AUC). The whole process was repeated 100 times to assess the variance of the model linked to the choice of the test dataset (Supplemental Fig. 2).

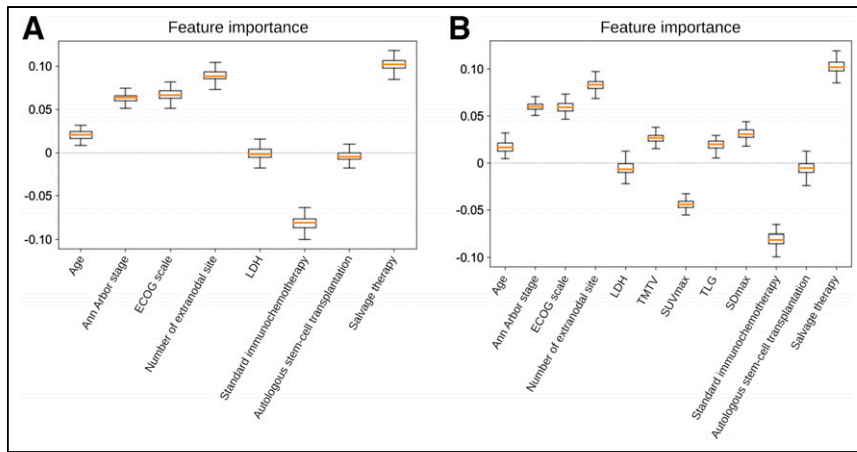
For the survival analysis, the model building was done in 2 main steps. A univariate analysis for each considered feature was conducted by fitting the Cox proportional-hazards model on the training dataset to select only those features significantly associated with the patient outcome using the training dataset. A penalized multivariate Cox analysis was then conducted using a 4-fold cross validation on the training dataset to select the best hyperparameters among the penalization strengths (Supplemental Table 2). A regression score (RS) was then derived for each patient by summing each term of the Cox proportional-hazards model (weight multiplied by the corresponding feature) (15). The population was then split according to the median of the RS to derive 2 survival groups. This RS cutoff was subsequently applied to the testing dataset. This process was repeated 100 times as for the previous 2-y PFS classification objective (Supplemental Fig. 3).

The 2 models (one for classification and one for survival) were implemented for 9 groups of different feature subsets, summarized in Table 1. Model 3, termed clinical, includes the same features as model 1, termed age-adjusted international prognostic index, with the difference being that the features are considered continuous in model 3 whereas a categorical score is used in model 1.

Since the consolidation arm was included in the features groups, none of the patients who died before the consolidation phase were considered in the study. Moreover, patients censored before 24 mo were not included for the 2-y PFS classification task.

### Patient Follow-up and Statistical Analysis

The 2-y PFS was defined as the time from randomization to disease progression, relapse, or death of any cause. A mean ROC curve for 2-y PFS and PET4 classification was generated taking into account the 100 loops. Accuracy, balanced accuracy, sensitivity, specificity, and AUC were computed. AUC among the 100 loops was compared using a 1-sided Wilcoxon



**FIGURE 2.** Feature importance according to relative weight attached to each feature in LR model for models 6 (A) and 7 (B) for 2-y PFS classification. ECOG = Eastern Cooperative Oncology Group; LDH = lactate dehydrogenase; SDmax = maximum distance between 2 lesions normalized by body surface area; TLG = total lesion glycolysis.

signed-rank test among the 9 models embedding different feature subsets. A similar approach was used to derive the most valuable ROI (among the 3 studied) to compute RFs. We reported the results of the 9 models using only the ROI leading to the best performance when radiomics is involved in the considered models. Feature importance was subsequently derived from the LR weights to highlight which features were the most informative. To this end, the best parameters of LR were chosen by 4-fold cross-validation on the training dataset, and then LR was retrained 100 times (Supplemental Fig. 2). Finally, box plots and sorted bar charts of absolute values were derived from LR weights.

Patient survival was estimated by Kaplan–Meier analysis, and survival curves were compared using the log-rank test. Box plots of

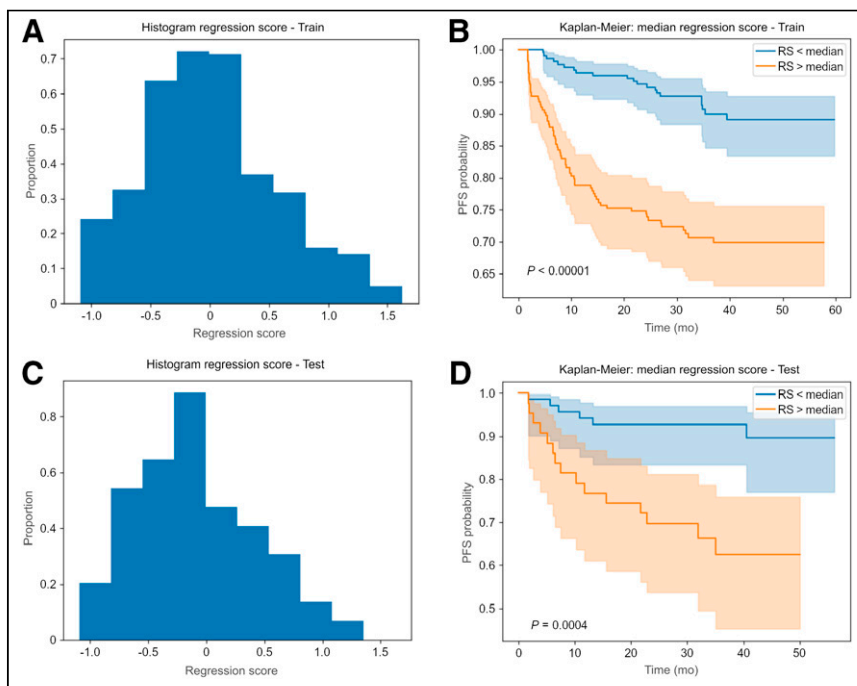
clinical characteristics similar to those of the entire GAINED population. Patient characteristics are summarized in Table 2. Among the 561 patients in our cohort, there were 386 early responders, 84 slow responders, and 91 who did not respond.

### 2-Year PFS Classification

Comparisons of the predictive values of radiomics extracted from different VOIs are presented in Supplemental Figure 4. The mean AUC was  $0.62 \pm 0.07$  for radiomics extracted from both the largest lesion and the TMTV but was  $0.54 \pm 0.07$  for radiomics extracted from the hottest lesion. Since the Wilcoxon test reported a significant difference ( $P < 0.0001$ ) between results extracted from either the largest lesion or TMTV and the hottest lesion, only radiomics computed from the largest lesion were considered.

ROC curves for models involving 1 group of features (from models 1 to 5) are shown in Figure 1A. The model with the highest predictive value was that including clinical features (model 3), with a mean AUC of  $0.65 \pm 0.07$ . Models involving consolidation features (model 2) and conventional PET (model 4) achieved a mean AUC of  $0.64 \pm 0.06$  and  $0.60 \pm 0.07$ , respectively, which are better than for model 1 (age-adjusted international prognostic index), which achieved a mean AUC of  $0.56 \pm 0.06$ . Only model 1 was significantly different from all others ( $P < 0.001$ ). Models combining several groups of features (from models 6 to 9) reached a mean AUC from  $0.72 \pm 0.06$  to  $0.69 \pm 0.06$  without significant differences in between.

The importance of each feature was derived from the weights resulting from the LR attached to each feature. A positive weight sign tends to classify patients in the positive class (progression occurred) when the feature value increases. Inversely, a negative weight sign tends to classify patients in



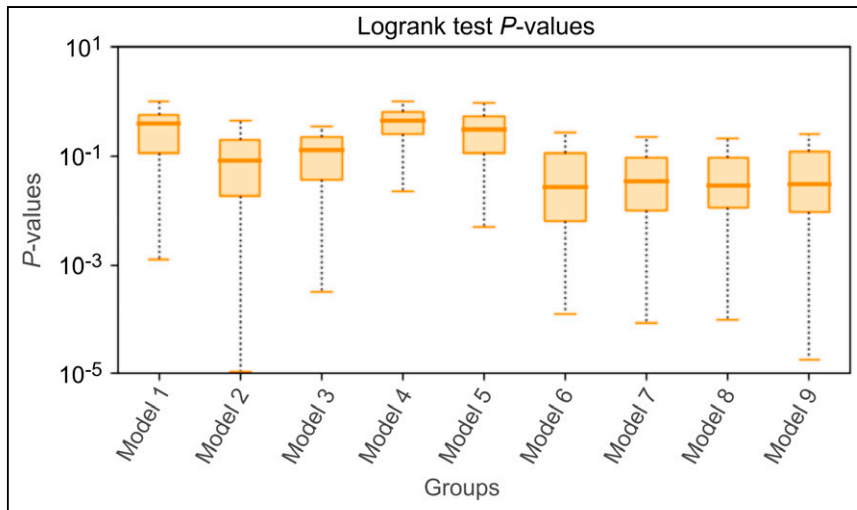
**FIGURE 3.** Example of RS for training set (A) and corresponding test set (C), along with associated Kaplan–Meier plots for PFS for training set (B) and test set (D). Low- and high-risk groups were dichotomized using median RS determined on training set and applied on test set.

$P$  values related to each model (over the 100 loops) were plotted and then were compared using a 1-sided Wilcoxon signed-rank test. The probability of selecting a feature after the univariate Cox analysis was also computed to highlight feature importance. All statistical analyses were done using Python version 3.7.9. A  $P$  value of less than 0.05 was considered statistically significant.

## RESULTS

### Patient Characteristics

Among the 670 patients enrolled in the GAINED study, 561 were considered for this analysis. For the 2-y PFS classification approach, the study included 545 patients, of whom 86 underwent an event before 24 mo. For the survival approach, the study included 561 patients, of whom 107 underwent an event before the last follow-up. These patients have demographic and



**FIGURE 4.** Log-rank *P* values over 100 loops for 9 models considered for survival analysis study.

the negative class when the feature value increases. The feature importance for models 6 and 7 is plotted in Figure 2. The number of extranodal sites, Eastern Cooperative Oncology Group scale, and Ann Arbor stage were among the most important features when considering clinical features. The consolidation treatment regimen received was also highly predictive of 2-y PFS. Adding conventional PET features (Fig. 2B) or radiomics did not notably change the ordering of the most important features (Supplemental Fig. 5).

#### Survival Analysis for PFS

The probability of selecting each feature for its superior significance compared with others (computed over the 100 loops) using the models combining several groups of features (from models 6 to 9) is presented in Supplemental Figure 6. The most important features selected in the previous classification task were also always selected for this current survival analysis study. An example relative to 1 loop of an RS histogram and associated Kaplan–Meier curves built using the median cutoff RS is shown in Figure 3 for model 6. The corresponding log-rank *P* values for the test dataset are presented in Figure 4. The median *P* value for models involving 1 group of features (from models 1 to 5) was lowest for the model involving only the consolidation treatment (model 2) and was significantly different

from models involving age-adjusted international prognostic index (model 1,  $P < 0.0001$ ), clinical parameters (model 3,  $P < 0.05$ ) conventional PET features (model 4,  $P < 0.0001$ ), or RFs (model 5,  $P < 0.0001$ ). Among models that combined the consolidation treatment plus one or more feature groups (from models 6 to 9), the median *P* value was the lowest for model 6 but without being significantly lower than for all other models (from models 7 to 9).

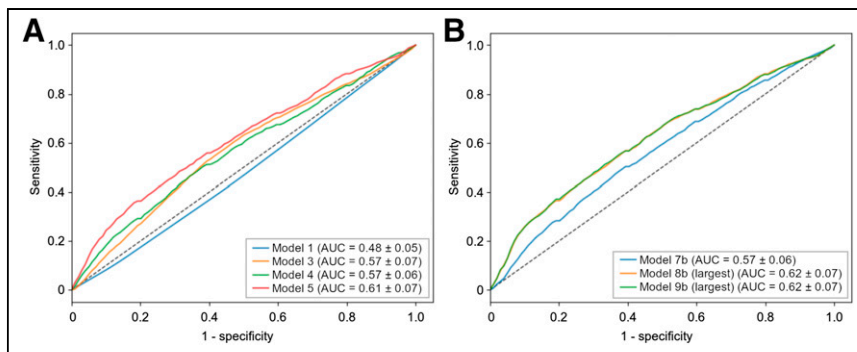
#### PET4 Prediction

Comparisons of predictive values of radiomics extracted from different VOIs are presented in Supplemental Figure 7. No significant difference was highlighted between the different VOIs. Then, for consistency reasons with the 2-y PFS classification, the largest one was retained.

ROC curves for models involving 1 group of features (models 1, 3, 4, and 5) are shown in Figure 5A. The model involving radiomics (model 5) was significantly better than all others ( $P < 0.0001$ ), with a mean AUC of  $0.61 \pm 0.07$ . When combined, only the models involving radiomics exhibited the best performance (Fig. 5B), with a mean AUC of  $0.62 \pm 0.07$  ( $P < 0.0001$ ).

#### DISCUSSION

The last 2 decades witnessed multiple trials exploring a PET-driven strategy to identify poorly responding DLBCL patients requiring more intensive salvage therapy. These works have yielded heterogeneous results, as many relied on qualitative visual evaluation (3). Assessing  $\Delta\text{SUV}_{\text{max}}$  after 2 and 4 cycles of induction regimens showed improved reproducibility and significantly reduced false-positive rates. This semiquantitative criterion was applied in 2 large multicentric prospective PET-guided trials (PETAL and GAINED). The first one, although disappointingly reporting no efficacy in escalation of PET2-positive patients to an intensified Burkitt chemotherapy, clearly defined  $\Delta\text{SUV}_{\text{max}}$  as a valid measure to differentiate patients with chemotherapy-sensitive tumors from those with chemotherapy-resistant tumors. In the GAINED study, as opposed to PETAL, positive interim PET's unfavorable prognostic value could be overcome with escalation of therapy and autologous stem cell transplantation. These data represented an important cornerstone in the management of DLBCL patients, suggesting that quantitative  $\Delta\text{SUV}_{\text{max}}$  criteria can be applied in routine practice to evaluate metabolic response and drive the therapeutic strategy. The question now arises as to what additional data would refine the prediction of early response to therapy. In a multiparametric approach, some teams are exploring circulating tumor DNA analysis, whereas some are investigating the value of PET biomarkers at baseline. Although the role of  $\text{SUV}_{\text{max}}$  at baseline to predict the 2-y PFS or overall survival was shown to be limited (4), evaluation of predictive value in DLBCL



**FIGURE 5.** Mean ROC curves for models 1, 3, 4, and 5 (A) and for models that combined several features (B) for PET4 prediction. Model 7b = combination of clinical and conventional PET features; model 8b = combination of clinical and radiomics; model 9b = combination of clinical and conventional PET features and radiomics.

patients using TMTV (5), dissemination (6), and RFs (9–17) has recently been considered.

In this study, our findings supported the fact that conventional PET metrics, dissemination, or RFs are predictive of outcome. Nevertheless, we extend these results by showing that these features seem to have no additional predictive capabilities compared with treatment arm or clinical features in the frame of the GAINED protocol. Interim PET assessment through the use of  $\Delta\text{SUV}_{\text{max}}$  allowed an accurate stratification of DLBCL patients into 3 risk groups, of which PET4-positive patients are those with the worst outcome, despite salvage therapy. The impact of the GAINED strategy in modifying conventional immunochemotherapy on the basis of these findings significantly and sufficiently improved patient outcome to override the value of the other PET parameters. The results were also suggested by the HOVON-84 trial, in which baseline TMTV did not add prognostic value to  $\Delta\text{SUV}_{\text{max}}$  (27).

This work also allowed exploration of the influence of VOI choice on the prognostic performance of radiomics. We were able to confirm the results of Eertink et al. (13) on interest in computing RFs from the TMTV or the largest lesion rather than the hottest lesion. Although this observation may seem surprising, it provides a better understanding of inpatient interlesional tumor heterogeneity. This is an important factor that is actually also explored by the  $\Delta\text{SUV}_{\text{max}}$  method. It compares the lesion with the highest  $\text{SUV}_{\text{max}}$ , at baseline and at interim PET, which is not necessarily the same hottest lesion as before the start of treatment, to measure the metabolic activity of the most active or aggressive tumor contingent.

Baseline PET characteristics showed limited performance in predicting results at PET4 and subsequently each patient's risk group and treatment arm. Nevertheless, unlike the 2-y PFS classification, radiomics seems to be of importance in that context since it outperformed the models that involved only the clinical information or the PET conventional features. This paves the way to the development of more advanced machine-learning models taking into account (or not) hand-crafted radiomics. Moreover, early identification of patients at highest risk could be of relevance in a context where innovative treatments, such as cell-based therapies, require a significant preparation and set-up phase.

## CONCLUSION

This study suggests that metrics extracted from  $^{18}\text{F}$ -FDG PET at baseline for this cohort of DLBCL patients and treated according to the PET-driven design of the GAINED protocol do not improve the 2-y PFS classification and PFS prediction. The model including the consolidation strategy, which was in turn directed by  $\Delta\text{SUV}_{\text{max}}$ , combined with clinical features achieved the best performances.

## DISCLOSURE

This work has been supported in part by ISITE NEXt (No. ANR-16-IDEX-0007), the French National Research Agency Labex IRON (No. ANR-11-LABX-0018-01), the INCa-DGOS-Inserm-12558 (SIRIC ILIAD), and the European Regional Development Fund, Pays-de-la-Loire (Connect Talent MILCOM), Nantes Métropole (Convention 2017-10470). Steven Le Gouill reports grants, personal fees, or nonfinancial support from Roche Genentech during the conduct of the study; personal fees from Celgene; and grants and personal fees from Janssen-Cilag, GILEAD/kite, and Servier outside the submitted work. René-Olivier Casanovas reports grants, personal fees, and nonfinancial support from Roche Genentech during the

conduct of the study; personal fees from MSD, BMS, Abbvie, Amgen, Celgene, Janssen, and Astra Zeneca; and grants and personal fees from Takeda and GILEAD/kite outside the submitted work. No other potential conflict of interest relevant to this article was reported.

## ACKNOWLEDGMENT

We warmly thank Enzo Lepoivre for his invaluable support in finalizing the manuscript.

## KEY POINTS

**QUESTION:** Do baseline  $^{18}\text{F}$ -FDG PET biomarkers add prognostic value to the PET-driven strategy of the GAINED study in DLBCL patients?

**PERTINENT FINDINGS:** Regardless of the VOI or features selection, conventional or radiomics  $^{18}\text{F}$ -FDG PET biomarkers extracted at baseline do not seem to improve the results in terms of 2-y PFS and survival prediction for patients treated according to the GAINED protocol.

**IMPLICATIONS FOR PATIENT CARE:** Interim PET assessment using  $\Delta\text{SUV}_{\text{max}}$  variation allows for accurate and simple monitoring and should be considered for use in routine practice in patients with advanced DLBCL.

## REFERENCES

1. Al Tabaa Y, Bailly C, Kanoun S. FDG-PET/CT in lymphoma: where do we go now? *Cancers (Basel)*. 2021;13:5222.
2. Sehn LH, Salles G. Diffuse large B-cell lymphoma. *N Engl J Med*. 2021;384:842–858.
3. Le Gouill S, Casanovas R-O. Interim PET-driven strategy in de novo diffuse large B-cell lymphoma: do we trust the driver? *Blood*. 2017;129:3059–3070.
4. Froot R, Burton C, Tsoumpas C, et al. Baseline PET/CT imaging parameters for prediction of treatment outcome in Hodgkin and diffuse large B cell lymphoma: a systematic review. *Eur J Nucl Med Mol Imaging*. 2021;48:3198–3220.
5. Ceriani L, Gritti G, Cascione L, et al. SAKK38/07 study: integration of baseline metabolic heterogeneity and metabolic tumor volume in DLBCL prognostic model. *Blood Adv*. 2020;4:1082–1092.
6. Cottreau A-S, Nioche C, Dirand A-S, et al.  $^{18}\text{F}$ -FDG PET dissemination features in diffuse large B-cell lymphoma are predictive of outcome. *J Nucl Med*. 2020;61:40–45.
7. Eertink JJ, Zwezerijnen GJC, Heymans MW, et al. Baseline PET radiomics outperforms the IPI risk score for prediction of outcome in diffuse large B-cell lymphoma. *Blood*. 2023;141:3055–3064.
8. Hatt M, Cheze Le Rest C, Antonorsi N, et al. Radiomics in PET/CT: current status and future AI-based evolutions. *Semin Nucl Med*. 2021;51:126–133.
9. Parvez A, Tau N, Hussey D, Maganti M, Metser U.  $^{18}\text{F}$ -FDG PET/CT metabolic tumor parameters and radiomics features in aggressive non-Hodgkin's lymphoma as predictors of treatment outcome and survival. *Ann Nucl Med*. 2018;32:410–416.
10. Aide N, Fruchart C, Nganoa C, Gac AC, Lasnon C. Baseline  $^{18}\text{F}$ -FDG PET radiomic features as predictors of 2-year event-free survival in diffuse large B cell lymphomas treated with immunochemotherapy. *Eur Radiol*. 2020;30:4623–4632.
11. Senjo H, Hirata K, Izumiyama K, et al. High metabolic heterogeneity on baseline  $^{18}\text{F}$ -FDG-PET/CT scan as a poor prognostic factor for newly diagnosed diffuse large B-cell lymphoma. *Blood Adv*. 2020;4:2286–2296.
12. Eertink JJ, van de Brug T, Wieggers SE, et al.  $^{18}\text{F}$ -FDG PET baseline radiomics features improve the prediction of treatment outcome in diffuse large B-cell lymphoma. *Eur J Nucl Med Mol Imaging*. 2022;49:932–942.
13. Eertink JJ, Zwezerijnen GJC, Cysouw MCF, et al. Comparing lesion and feature selections to predict progression in newly diagnosed DLBCL patients with FDG PET/CT radiomics features. *Eur J Nucl Med Mol Imaging*. 2022;49:4642–4651.
14. Lue KH, Wu YF, Lin HH, et al. Prognostic value of baseline radiomic features of  $^{18}\text{F}$ -FDG PET in patients with diffuse large B-cell lymphoma. *Diagnostics (Basel)*. 2020;11:36.
15. Ceriani L, Milan L, Cascione L, et al. Generation and validation of a PET radiomics model that predicts survival in diffuse large B cell lymphoma treated with R-CHOP14: a SAKK 38/07 trial post-hoc analysis. *Hematol Oncol*. 2022;40:11–21.

16. Zhang X, Chen L, Jiang H, et al. A novel analytic approach for outcome prediction in diffuse large B-cell lymphoma by [<sup>18</sup>F]FDG PET/CT. *Eur J Nucl Med Mol Imaging*. 2022;49:1298–1310.
17. Jiang C, Li A, Tend Y, et al. 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*. 2022;49:2902–2916.
18. Le Gouill S, Ghesquières H, Oberic L, et al. Obinutuzumab vs rituximab for advanced DLBCL: a PET-guided and randomized phase 3 study by LYSA. *Blood*. 2021;137:2307–2320.
19. Barrington SF, Zwezerijnen BGJC, de Vet HCW, et al. Automated segmentation of baseline metabolic total tumor burden in diffuse large B-cell lymphoma: which method is most successful? A study on behalf of the PETRA consortium. *J Nucl Med*. 2021;62:332–337.
20. van Griethuysen JJM, Fedorov A, Parmar C, et al. Computational radiomics system to decode the radiographic phenotype. *Cancer Res*. 2017;77:e104–e107.
21. Bailly C, Bodet-Milin C, Couespel S, et al. Revisiting the robustness of PET-based textural features in the context of multi-centric trials. *PLoS One*. 2016;11:e0159984.
22. Desseroit MC, Tixier F, Weber WA, et al. Reliability of PET/CT shape and heterogeneity features in functional and morphologic components of non-small cell lung cancer tumors: a repeatability analysis in a prospective multicenter cohort. *J Nucl Med*. 2017;58:406–411.
23. van Velden FHP, Kramer GM, Frings V, et al. Repeatability of radiomic features in non-small-cell lung cancer [<sup>18</sup>F]FDG-PET/CT studies: impact of reconstruction and delineation. *Mol Imaging Biol*. 2016;18:788–795.
24. Pfähler E, Mesotten L, Zhovannik I, et al. Plausibility and redundancy analysis to select FDG-PET textural features in non-small cell lung cancer. *Med Phys*. 2021;48:1226–1238.
25. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. *J Artif Intell Res*. 2002;16:321–357.
26. He H, Garcia EA. Learning from imbalanced data. *IEEE Trans Knowl Data Eng*. 2009;21:1263–1284.
27. Burggraaff CN, Eertink JJ, Lugtenburg PJ, et al.; HOVON imaging working group and the HOVON Lymphoma Working Group. <sup>18</sup>F-FDG PET improves baseline clinical predictors of response in diffuse large B-cell lymphoma: the HOVON-84 study. *J Nucl Med*. 2022;63:1001–1007.