Translational Oncology 14 (2021) 101188

Contents lists available at ScienceDirect

Translational Oncology

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

CT-based radiomics model with machine learning for predicting primary treatment failure in diffuse large B-cell Lymphoma

Raoul Santiago [a,b,†], Johanna Ortiz Jimenez [a,‡], Reza Forghani [b,c,d,e,*], Nikesh Muthukrishnan [b,c], Olivier Del Corpo [e], Shairabi Karthigesu [e], Muhammad Yahya Haider [e], Caroline Reinhold [c,e], Sarit Assouline [a,b]

[a] Jewish General Hospital - McGill University, Canada
[b] Segal Cancer Centre and Lady Davis Institute for Medical Research, Canada
[c] Augmented Intelligence & Precision Health Laboratory (AIPHL) of the Department of Radiology and the Research Institute of McGill University Health Centre, Canada
[d] Gerald Bronfman Department of Oncology, Canada
[e] McGill University, Canada

ARTICLE INFO

Keywords:
Radiomics
Diffuse Large B-cell Lymphoma
Refractory
Quantitative imaging
Biomarkers

ABSTRACT

Biomarkers which can identify Diffuse Large B-Cell Lymphoma (DLBCL) likely to be refractory to first-line therapy are essential for selecting this population prior to therapy initiation to offer alternate therapeutic options that can improve prognosis. We tested the ability of a CT-based radiomics approach with machine learning to predict Primary Treatment Failure (PTF)-DLBCL from initial imaging evaluation. Twenty-six refractory patients were matched to 26 non-refractory patients, yielding 180 lymph nodes for analysis. Manual 3D delineation of the total node volume was performed by two independent readers to test the reproducibility. Then, 1218 handcrafted radiomic features were extracted. The Random Forests machine learning approach was used as a classifier for constructing the prediction models. Seventy percent of the nodes were randomly assigned to a training set and the remaining 30% were assigned to an independent test set. The final model was tested on the dataset from the 2 readers, showing a mean accuracy, sensitivity and specificity of 73%, 62% and 82%, respectively, for distinguishing between refractory and non-refractory patients. The area under the receiver operating characteristic curve (AUC) was 0.83 and 0.79 for the two readers. We conclude that machine learning CT-based radiomics analysis is able to identify a priori PTF-DLBCL with a good accuracy.

Introduction

Over 60% of patients with Diffuse Large B Cell Lymphoma (DLBCL) are cured after first-line R-CHOP (rituximab, cyclophosphamide, doxorubicin, vincristine and prednisone) [1]. However, up to 15% experience Primary Treatment Failure (PTF) with a median survival not exceeding one year [2]. Emerging therapies, such as chimeric antigen receptor T-cell, produce high response rates in relapsed/refractory DLBCL [3] and could benefit patients at risk for PTF to R-CHOP. Ideally, these patients should be identified prior to receiving first-line therapy.

Determination of double-hit status allows for the selection of approximately 5% of newly diagnosed DLBCL patients unlikely to derive meaningful benefit from R-CHOP [4]. The revised international prognostic index (R-IPI) and presence of TP53 mutation predict long-term survival but cannot identify patients likely to experience PTF [5,6]. Also, diagnostic imaging is routinely used for staging purposes, but no conventional radiological findings have been correlated with PTF-DLBCL [7].

Radiomics is the high-throughput extraction of large amounts of image features from radiographic images and mining and use of this data for improved clinical decision support [8–11]. Radiomics can assist in diagnosis, prognosis, staging and in predicting treatment response in oncology [12–14].

Quantitative analysis of initial and interim 18FDG-PET/CT (PET-scan) has been studied as a prognostic marker in DLBCL [15–20]. Combining the texture analysis of PET-scans to functional parameters

* Corresponding author.
E-mail address: reza.forghani@mcgill.ca (R. Forghani).
† Present address for Raoul Santiago: CHU de Québec – Université Laval, Canada
‡ Present address for Johanna Ortiz Jimenez: Kingston General Hospital - Queen's University, Canada

https://doi.org/10.1016/j.tranon.2021.101188
Received 18 April 2021; Received in revised form 28 June 2021; Accepted 23 July 2021

correlates better with survival than Standardized Uptake Values (SUV) analysis alone [16–18]. However, data on the potential value of diagnostic CT-scan extracted quantitative features in lymphoma are sparse but could be of added value in predicting outcomes [7,13,21].

Identifying a reliable biomarker predicting PTF-DLBCL would allow treatment optimization, thereby sparing exposure to ineffective treatment, which can result in morbidity and resistance to subsequent therapies. Thus, we aimed to develop a machine-learning CT-based radiomics model for PTF-DLBCL prediction.

Materials and Methods

Patients, CT scans, and nodes selection

This single-center study included adult patients diagnosed with de novo DLBCL from 2009 to 2018, who received R-CHOP or similar frontline therapy (Table S1). Patients who received previous chemotherapy or radiotherapy, for instance for prior indolent lymphoma, were excluded. This study was conducted in accordance with the Declaration of Helsinki and was approved by the Research Ethic Committee of the CIUSSS of West-Central Montreal with a waiver for patient informed consent in the context of a retrospective study (ethical approval code: 2019-1271).

A CT-scan at diagnosis, prior to treatment initiation, was required for image acquisition to be included in the prediction model. Venous phase CT-scans of the chest and neck, and portal venous phase CT-scans of the abdomen and pelvis were analyzed. A whole-body CT scan was obtained as per standard institutional protocol after the administration of 80 mL of iopamidol (Isovue 300; Bracco Diagnostics, Princeton, New Jersey) injected at a rate of 2 mL/s and scans were acquired at a kVp of 120 with a 40 mm beam collimation, a 0.8–second rotation time, a 0.984:1 helical pitch, and a field of view 50 cm (chest; abdomen and pelvis) or 32 cm (neck). Images were reconstructed into 2.5 mm sections (neck), 3 mm sections (chest), or 5 mm (abdomen and pelvis). We retrospectively reviewed the patients diagnosed for DBLCL in our center on a 10-year period, from 2009 to 2018, who were eligible to this study. Prior to analysis, the patients were divided in two groups for comparison: PTF and non-PTF DBLCL. Refractory patients, assigned to the PTF group, were defined by progression of disease (PD) during R-CHOP, or failure to achieve a complete response (CR) after at least 4 cycles, as per the Lugano criteria [22]. In the latter group, non-refractory patients (CR without relapse within 6 months of therapy) were matched 1:1 on sex and R-IPI (including age, stage, performance status, LDH level and extra-nodal sites) [5]. When multiple non-refractory patients could be matched with a refractory patient, the one with the closest age was favored.

On the CT-scan for initial diagnosis, the lymph nodes measuring $\geq$1.5 cm in greatest diameter were evaluated. To limit the number of nodes selected per patient, a maximum of 6 nodes at each of 4 nodal sites (abdomen, chest, axilla and neck) were included per patient to mirror the Lugano response criteria (standard response criteria in lymphoma). When the number of eligible nodes in a specific nodal site exceeded six, only the six nodes with the largest maximum diameter in any axis were kept for analysis. The nodes were categorized in response categories: Refractory Node (RN), Partial Response (PR) and CR, as per the Lugano criteria [22]. The node selection and categorization into response groups were assessed by the two readers (JOJ and RS) in order to reach a consensus. A double validation for node selection was chosen to ensure the respect of inclusion criteria and to avoid misclassification. When a consensus could not be reached either for node eligibility or for group attribution, the decision was left to the senior radiologist (RF).

Contouring and feature extraction

The open-source 3D Slicer software v4.10.2 [23] was used for the volume of interest (VOI) delineation. Each node was manually contoured by two independent readers to obtain a duplicated dataset from each independent reader in order to test the reproducibility of the model. When a conglomerate of nodes was present, the conglomerate was included as a single VOI unless a cleavage line could be identified in between nodes therefore delimiting two or more distinct VOI. To avoid inclusion of extra-nodal material, the contouring was reviewed by an experienced senior oncologic radiologist (RF) and, if needed, the VOI was edited by the initial reader accordingly. The central revision aimed to ensure the respect of the inclusion/exclusion criteria for node selection and was not meant to flatten the inter-individual contouring variation. Lymph nodes altered by artifact and lymph nodes with central necrosis or cystic nodal change (area of low attenuation, between -10 and 30 Hounsfield units, and greatest diameter $\geq$ 3 mm) were labeled by each reader. The nodes severely obscured by artifact were excluded. The nodes with necrosis or cystic changes, referred as subjective necrosis, were kept in the analysis and included as a variable in the construction of the model.

A total of 1218 features were extracted from VOI using PyRadiomics open source software [24]. Features were extracted using the following libraries and versions: PyRadiomics 2.2.0, Numpy 1.13.1, SimpleITK 1.1.0, PyWavelet 1.0.0 and Python 2.7.13. The extracted features included both low and high abstract level features: 18 first-order statistic for pixel level, 14 3D shape-based as well as 68 gray level texture features, totaling 100 features. In addition to the default PyRadiomics features, 13 multiple spatial filters (5 Laplacian of Gaussian and 8 wavelet filters) were applied to the image to extract fine and coarse features. Each feature, at the exception of shape-based features, were extracted from filtered images (86 features with 13 filters = 1118 additional features). A table is available in the supplemental material summarizing the list of features and filters used (Table S2). Along with the radiomics features, the nodal site and subjective assessment of necrosis were also evaluated as additional features.

Statistical analysis

Initially, multiple t-tests were used to assess the differential expression of each single feature between CR and non-CR nodes. The adjusted p value was carried out by the Benjamini Yekutieli method. A hierarchical clustering by Euclidean distance with dendrogram was generated using the enhanced heatmap function, heatmap.2, in gplots package (v.3.1.1) for R.

The Random Forests classifier was trained to determine refractory and non-refractory patients by comparing the non-CR nodes from refractory patients to CR nodes from non-refractory patients from the first reader's dataset.. For an unbiased assessment of the model accuracy, the data was first split into training and test dataset. 30% of the selected nodes were randomly selected and set aside as the independent test (prediction) group [11,13]. The remaining 70% were used to train, tune and validate the prediction model. Note, the same nodes were initially allocated to training and test dataset for both readers for a fair comparison, however, since each reviewer evaluated the lymph nodes independently, there is a small discrepancy based on reviewer variation for exclusion of nodes they deemed as degraded by artifact. As a result, there is a small variation between cases used to train and evaluate reader 1 and reader 2 data. Due to the abundance of features, we first eliminated features that were found to be highly correlated on the training set. Numerous correlation cut-offs were evaluated and 0.6 was manually selected as it removed highly correlated features and maximized the number of features available during recursive feature elimination without burdening the machine severely (through exhaustive computations).We performed recursive feature elimination using the rfeControl (caret package) to identify the top performing features for a random forest (optimizing for accuracy). Repeated 10-fold cross validation was implemented to evaluate the top performing features on the training set (training data was split into 10 folds, model was trained on the first nine and evaluated on the 10^{th}. After repeating 10 times, the average

performance was recorded and compared). Furthermore, with the top performing features, the RF (randomForest package) was tuned for number of trees (using a grid search between 5 to 20,000 trees) by evaluating the model's accuracy with repeated cross-validation on the training set. The top performing features and tuned number of trees were then used to train a final RF model and evaluate the test set. This process was repeated several times across the same five seeds for both readers and the average performance was recorded and compared. The list of the packages used for this study and their version are available in the supplemental material (Table S3). The final model, built on the dataset from reader 1, was applied to the dataset from reader 2 to test the model reproducibility and generalizability of the selected features. Each independent node was considered to predict the patient's response to treatment.

Statistical analyses were carried out with GraphPad Prism v8.3.1 and R software v3.6.1 and the R package random Forest v4.6–14 for machine learning.

Results

Patients' characteristic

Twenty-six refractory patients were identified with 75 nodes which did not reach a CR: 55 RN and 20 PR nodes. The refractory patients were matched to 26 non-refractory patients, totaling 105 CR nodes. Eight (6 CR and 2 non-CR) and 10 (8 CR and 2 non-CR) nodes with significant artifact were excluded for reader 1 and 2, respectively. The median age at diagnosis was 65 and 71 years for refractory and non-refractory patients, respectively. The two groups were matched on sex and R-IPI, with a sex-ratio M/F of 1.36 and a median R-IPI of 3 (69% of R-IPI $\geq$ 3). In the refractory and non-refractory group, 96% and 88% of patients had a stage $\geq$ 3 and 31% and 19% had an Eastern Cooperative Oncology Group (ECOG) performance status $\geq$ 2 at diagnosis, respectively (Table 1). Three of 18 tested refractory patients and 0 of 11 tested non-refractory patients had double-hit DLBCL (the screening for double-hit DLBCL was following a local algorithm based on cell of origin, proliferation rate and MYC/BCL2 double-expresser) [25].

Heat map signature

We first explored a basic statistical association between non-CR and CR nodes. The multiple t-test found 35 features with an absolute log2 fold change $\geq$1 and adjusted p value < 0.05. The fig. 1 represents a heat map with the hierarchy clustering by Euclidean distance analysis for these features. The signature highlights a loose clustering between non-CR and CR nodes but did not identify distinct phenotypes.

Machine learning model

We then trained a machine learning classifier using Random Forests (RF) for the classification of non-CR vs CR nodes. During an initial step of hyper-parameter reduction and identification of highly correlated features, the 1218 radiomic features were reduced to 66, which were combined with 2 additional features: nodal site and subjective necrosis. Ten features were selected for the final model (Table S4), all of which had a good correlation between readers (Spearman test with median r = 0.8 and all adjusted $p$values < 0.0001). Neither nodal site nor subjective necrosis was retained in the final model.

The final model was trained on the dataset from the first reader and then applied on the dataset from the second reader. In the independent test set for the final model, the mean accuracy, sensitivity and specificity for distinguishing between refractory and non-refractory patients were 73%, 62% and 82%, respectively. The mean predictive positive and negative values were 0.77 and 0.71, respectively. Table 2 shows the performance of the independent test set by reader for the final model.

Table 1
Patient characteristics at initial diagnosis of diffuse large B-cell lymphoma (DLBCL)

Clinical features	Refractory patients (n = 26)	Matched non-refractory patients (n = 26)
Age (year) *	65 (24–86)	71 (28–87)
Sex ratio (M/F)	1.36	1.36
Stage $\geq$ 3	25 (96%)	23 (88%)
ECOG $\geq$ 2	8 (31%)	5 (19%)
R-IPI $\geq$ 3	18 (69%)	18 (69%)
Double hit DLBCL	3/18 **	0/11 **

* median (range)
** denominator is the number of patients for whom double hit status has been assessed

Area under the ROC curve and comparison to qualitative imaging

The area under the receiver operating characteristic (ROC) curve (AUC) was 0.83 (CI 95%: 0.81–0.84) and 0.79 (95% CI: 0.76–0.81) for reader 1 and 2, respectively. Fig. 2 demonstrates an AUC for the final model significantly superior to that of the semantic radiology marker of subjective necrosis (AUC = 0.56 (95% CI:0.49–0.63) and 0.52 (95% IC:0.45–0.59), for reader 1 and 2, respectively).

To illustrate ability to distinguish refractory from non-refractory disease, Fig. 3 compares a refractory and a responsive node that appear visually similar but was accurately distinguished by our model.

Altogether, this suggests that a CT-based radiomics model is able to classify refractory and non-refractory nodes with an accuracy that is currently unmatched with conventional imaging evaluation.

Discussion

In this study, we demonstrated that a machine learning model, analyzing the quantitative imaging parameters derived from diagnostic CT-scans of DLBCL patients, provides a good accuracy for discriminating the patients likely to be refractory to standard therapy.

The relationship of each independent feature to the outcome, CR or non-CR nodes, assessed by multiple t-test, revealed 35 features significantly different between the 2 groups (absolute log2 fold change $\geq$1 and adj $p$value < 0.05). The hierarchy clustering heat map built with these 35 discriminative features highlighted distinct signatures where refractory and non-refractory nodes loosely clustered together (Fig. 1). However, some of the features were highly correlated. It has been shown that a selection method based on independent feature characteristics without considering the concordant and discordant effect between groups of features can limit the classification performance [11].

A multivariate classifier with machine learning method was then used to build a model for the prediction of PTF-DLBCL. The prediction model was built with a randomly selected training set representing 70% of the nodes. An internal evaluation was performed with the test set from the 30% remaining nodes, providing an estimate of the unbiased generalization error. The test set from the final machine learning model performed with a specificity exceeding 80% for the 2 readers. In the context of a prediction test that could eventually be clinically implemented for patient selection and treatment decision, it is critical to develop a test with high specificity. The mean accuracy was 73%, with a mean sensitivity of 62%.

Our model showed a mean AUC of 0.81 for predicting refractory disease. In comparison, an interim PET-scan based study showed an AUC of 0.677 for response prediction [19]. Other studies showed that initial 18FDG-PET/CT scan can help predict long term survival, however they were not designed to assess for PTF-DLBCL [16,17].

Intratumor heterogeneity has been shown to be a strong imaging predictor of poor prognosis in solid tumors and in DLBCL [14,16–18]. Consistently, two of the final features in our model relate to

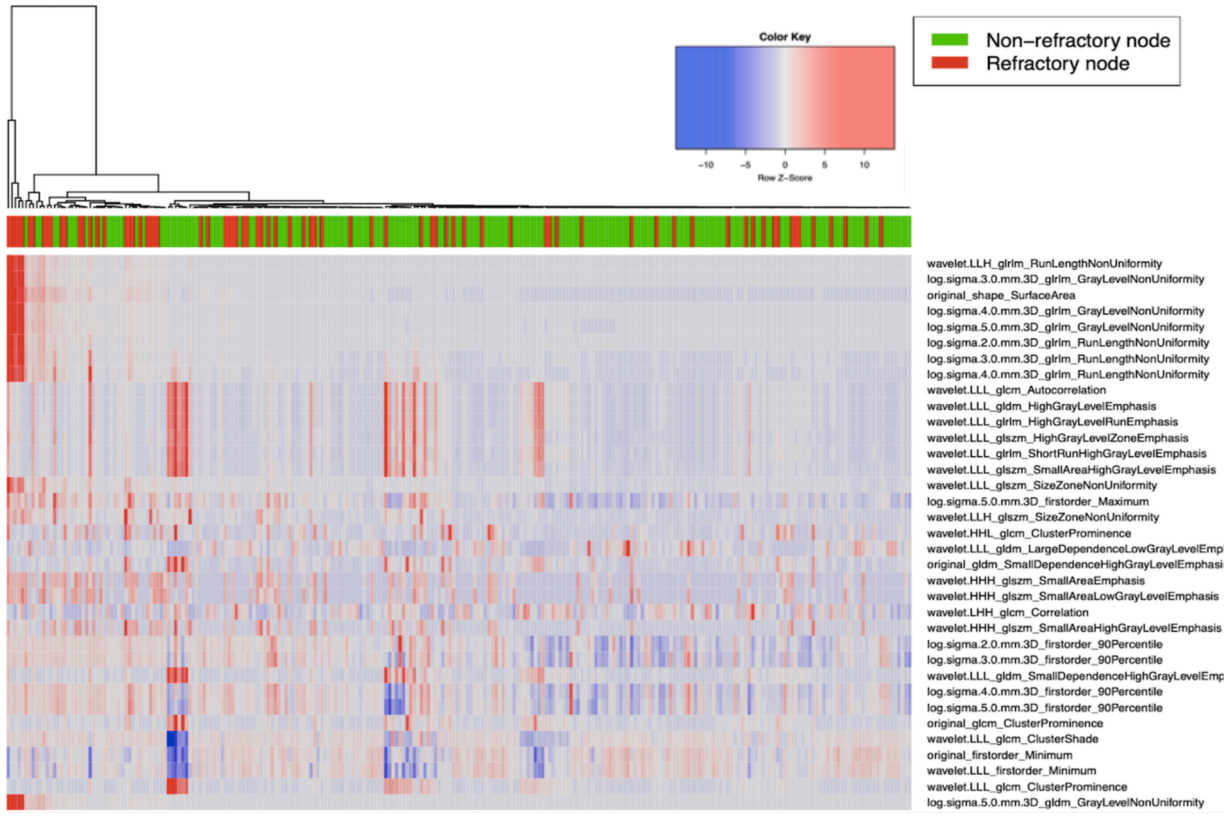

Fig.1. Heat map with hierarchy clustering by dendrogram analysis per node for the 35 most significant 3D radiomics features with absolute log2 change fold ≥1 and q.value < 0.05.

Table 2
Performance of the final testing radiomics model for each independent reader.

Reader	Sensitivity	Specificity	PPV	NPV	Accuracy	AUC
Reader 1 (95% CI)	0.66 (0.57–0.74)	0.84 (0.74–0.93)	0.77 (0.75–0.80)	0.73 (0.69–0.77)	0.76 (0.74–0.79)	0.83 (0.81–0.84)
Reader 2 (95% CI)	0.59 (0.49–0.69)	0.81 (0.76–0.85)	0.76 (0.67–0.86)	0.69 (0.66–0.73)	0.70 (0.68–0.73)	0.79 (0.76–0.83)

95% CI: 95% confidence interval, *AUC*: Area under the receiver operating characteristic (ROC) curve, *NPV*: negative predictive value, *PPV*: positive predictive value.

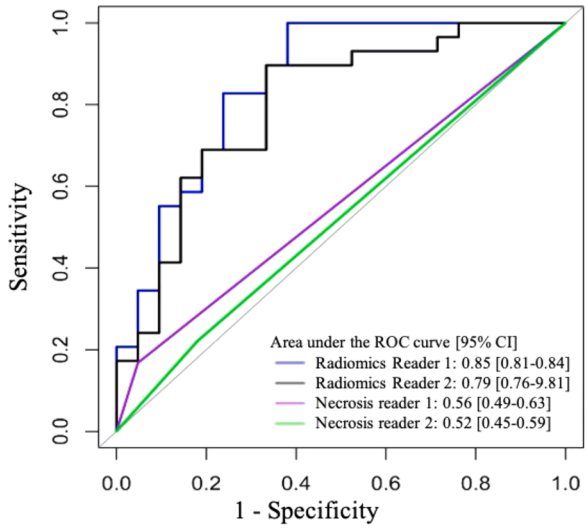

Fig. 2. Area under the receiver operating characteristic (ROC) curves (AUC) per reader comparing the final testing radiomics model to a qualitative/semantic radiology marker, the presence of subjective necrosis.
95% CI: 95% confidence interval, *ROC*: receiver operating characteristic

non-uniformity, thus identifying tumor heterogeneity as a prognostic marker as well. The prediction ability of texture features, such as intratumor heterogeneity, encourages to further explore contrast-enhanced CT-scan derived images in future radiomics models. The features could be extracted either from CT-Scan or from contrast-enhanced 18FDG-PET/CT.

Inter-observer variability from manual contouring is a limitation that has been frequently pointed out for radiomics [8,12]. Our study has been designed with double contouring by 2 independent readers to reduce the contouring bias. We observed a high inter-reader correlation for the 10 features selected in the final model (Spearman adjusted *p*value < 0.0001 for the all the final features). As an additional test for reproducibility, the model was built with data from reader 1 and applied to reader 2. As shown in Table 1 and in Fig. 2, the final model resulted in very similar performance between the two readers with confidence interval overcrossing for every test performance characteristic. Manual segmentation is also time consuming, making it less applicable to clinical implementation in "real-life" as a biomarker [12]. New approaches of segmentation with semi or full-automation algorithms should be tested in the future to accelerate the process and enable for its routine use in larger population.

Beside the extent of disease used for staging, there is no validated radiology method able to predict the outcome of DLBCL at diagnosis [7, 22]. The presence of tumor necrosis has been described as a prognostic

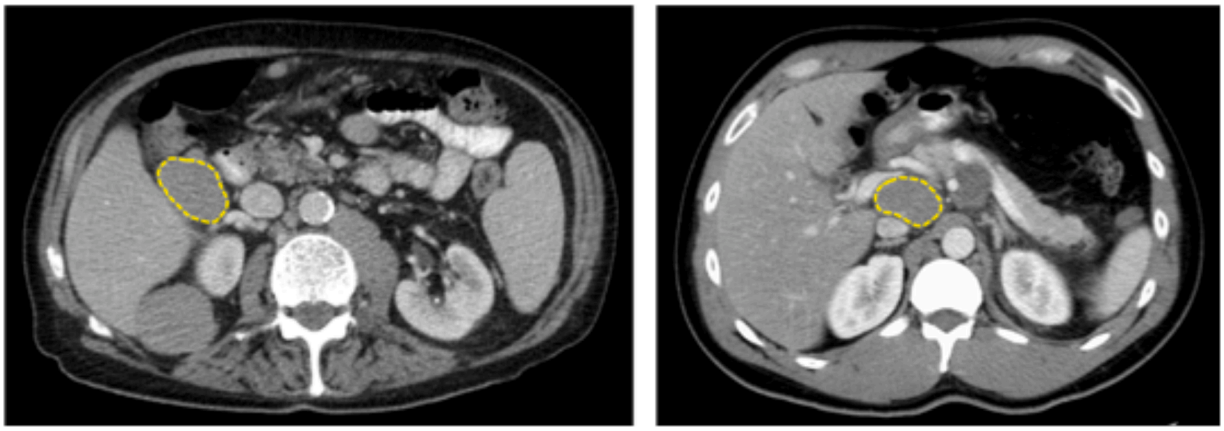

Fig. 3. Comparison of 2 nodes from patients CT-scan. The left picture shows a responsive node (achieved a complete response) from a non-refractory patient and the right picture a refractory node (did not achieve a complete response) from a refractory patient. The two nodes have been successfully distinguished by the radiomics model despite a very similar visual appearance.

marker for DLBCL [7,20]. As shown on the ROC curves (Fig. 2), our radiomics model performs better to discriminate PTF-DLBCL, compared to this qualitative radiology marker. Also, Fig. 3 compares two intra-abdominal nodes that could not visually be distinguished but were accurately labeled as refractory and non-refractory by our model. This strongly suggests that quantitative image-based features can predict PTF-DLBCL with an accuracy that is unequaled by conventional image interpretation of semantic features.

However, our study is limited by the small number of patients. DLBCL is characterized by a wide heterogeneity that can be subdivided within up to five distinct molecular subgroups [26,27]. The limited number of patients included in this study might not represent this molecular variability. To flatten the differences within our two comparative groups, refractory and non-refractory, the patients have been matched on the most significant clinical outcome predictor, the R-IPI score. Beyond inter-individual variability, the coexistence of different clonal population within a single patient's disease is acknowledged in a various number of cancers. Recently, the same spatial diversity was highlighted in DLBCL with frequent mutational disparity between different tumor sites in a same patient, even at initial diagnosis prior to treatment initiation [28]. This phenomenon can explain mixed response in some DLBCL patients where one or more lesions are progressing on treatment while others are responding. The existence of mixed response will still categorize the patient as refractory [29]. To take into account the intra-individual heterogeneity, each node was considered independently to test patient's response to treatment. We believe that a radiomics approach has the potential to better represent the complexity of a patient's disease than a single biopsy that might underestimate the spatial heterogeneity.

Our study showed that CT-based quantitative image-based features can predict PTF-DLBCL with a high accuracy and with good inter-reader reproducibility. There is no other known imaging model, either quantitative or qualitative, offering such an efficient prediction model. Future studies are planned to expand the dataset to develop more powerful algorithms that would improve the performance, and to validate the model on an external independent dataset.

Conclusions

This study suggests that CT-based radiomics with machine learning classifier might be able to discriminate PTF-DLBCL from non-refractory disease at the time of diagnosis and prior to therapy initiation. Our CT-based radiomics model analyzing nodal texture could discriminate refractory from non-refractory patients. The contrast enhanced CT used for our study is the favored method, along with PET CT scan, to assess disease extent and response to therapy in the context of clinical trials [22]. This quantitative contrast-enhanced CT scan-based model could be assessed in combination with metabolic PET-scan biomarkers. Our study is limited by a small dataset that might not depict the large molecular heterogeneity of this disease but the nodes independently increased our dataset and helped overcome intra-individual spatial heterogeneity of the tumor. The good inter-observer reliability suggests this approach to be reproducible but that needs to be confirmed on an external evaluation (test) set and a larger dataset is needed to confirm the generalizability of the findings. Future work in this area will focus on automating the identification of nodal structures, analysis of non-nodal involvement by DLBCL, and possibly combining SUV data from the PET scan.

CRediT authorship contribution statement

Raoul Santiago: Conceptualization, Methodology, Data curation, Formal analysis, Writing – original draft, Visualization. **Johanna Ortiz Jimenez:** Conceptualization, Methodology, Data curation, Formal analysis. **Reza Forghani:** Conceptualization, Writing – review & editing, Supervision, Project administration, Resources, Funding acquisition. **Nikesh Muthukrishnan:** . **Olivier Del Corpo:** Data curation. **Shairabi Karthigesu:** Data curation. **Muhammad Yahya Haider:** Data curation. **Caroline Reinhold:** Conceptualization, Resources. **Sarit Assouline:** Conceptualization, Writing – review & editing, Supervision, Project administration, Funding acquisition.

Declaration of Competing Interest

R.F. has acted as a consultant and speaker and has had research agreements with GE Healthcare. R.F. is a founder and stockholder of 4Intelligent Inc. and a founding member and president of Montreal Imaging Experts Inc..

Funding

This research was funded by Fonds de recherche du Québec - Santé (FRQS) (R.F.), Fondation de l'association des radiologistes du Québec (FARQ) (R.F.) and the DIVCO Foundation (S.A.).

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.tranon.2021.101188.

References

[1] P. Feugier, et al., 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 (2005) 4117–4126, https://doi.org/10.1200/JCO.2005.09.131.

[2] M. Crump, et al., Outcomes in refractory diffuse large B-cell lymphoma: results from the international SCHOLAR-1 study, Blood 130 (2017) 1800–1808, https://doi.org/10.1182/blood-2017-03-769620.

[3] C. Gisselbrecht, E. Van Den Neste, How I manage patients with relapsed/refractory diffuse large B cell lymphoma, Br J Haematol 182 (2018) 633–643, https://doi.org/10.1111/bjh.15412.

[4] N.A. Johnson, et al., Lymphomas with concurrent BCL2 and MYC translocations: the critical factors associated with survival, Blood 114 (2009) 2273–2279, https://doi.org/10.1182/blood-2009-03-212191, doi:blood-2009-03-212191 [pii].

[5] L.H. Sehn, et al., 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 109 (2007) 1857–1861, https://doi.org/10.1182/blood-2006-08-038257.

[6] Z.Y. Xu-Monette, et al., Mutational profile and prognostic significance of TP53 in diffuse large B-cell lymphoma patients treated with R-CHOP: report from an International DLBCL Rituximab-CHOP Consortium Program Study, Blood 120 (2012) 3986–3996, https://doi.org/10.1182/blood-2012-05-433334.

[7] H.J.A. Adams, et al., Prognostic value of tumor necrosis at CT in diffuse large B-cell lymphoma, Eur J Radiol 84 (2015) 372–377, https://doi.org/10.1016/j.ejrad.2014.12.009.

[8] R.J. Gillies, P.E. Kinahan, H. Hricak, Radiomics: images are more than pictures, they are data, Radiology 278 (2016) 563–577, https://doi.org/10.1148/radiol.2015151169.

[9] V. Kumar, et al., Radiomics: the process and the challenges, Magn Reson Imaging 30 (2012) 1234–1248, https://doi.org/10.1016/j.mri.2012.06.010.

[10] P. Lambin, et al., Radiomics: extracting more information from medical images using advanced feature analysis, Eur J Cancer 48 (2012) 441–446, https://doi.org/10.1016/j.ejca.2011.11.036.

[11] P. Savadjiev, et al., Image-based biomarkers for solid tumor quantification, Eur Radiol 29 (2019) 5431–5440, https://doi.org/10.1007/s00330-019-06169-w.

[12] R. Forghani, et al., Radiomics and artificial intelligence for biomarker and prediction model development in oncology, Comput Struct Biotechnol J 17 (2019) 995–1008, https://doi.org/10.1016/j.csbj.2019.07.001.

[13] M. Seidler, et al., Dual-Energy CT texture analysis with machine learning for the evaluation and characterization of Cervical Lymphadenopathy, Comput Struct Biotechnol J 17 (2019) 1009–1015, https://doi.org/10.1016/j.csbj.2019.07.004.

[14] H.J. Aerts, et al., Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach, Nat Commun 5 (2014) 4006, https://doi.org/10.1038/ncomms5006.

[15] H. Schoder, et al., Prognostic Value of Interim FDG-PET in diffuse large cell lymphoma: results from the CALGB 50303 clinical trial, Blood (2020), https://doi.org/10.1182/blood.2019003277.

[16] L. Ceriani, et al., SAKK38/07 study: integration of baseline metabolic heterogeneity and metabolic tumor volume in DLBCL prognostic model, Blood Adv 4 (2020) 1082–1092, https://doi.org/10.1182/bloodadvances.2019001201.

[17] H. Senjo, et al., High metabolic heterogeneity on baseline 18FDG-PET/CT scan as a poor prognostic factor for newly diagnosed diffuse large B-cell lymphoma, Blood Adv 4 (2020) 2286–2296, https://doi.org/10.1182/bloodadvances.2020001816.

[18] B. Ferrer Lores, et al., Prognostic value of radiomics signature by diagnostic 18F-FDG PET/CT analysis in aggressive Non-Hodgkin's Lymphoma, Blood 132 (2018) 1703, https://doi.org/10.1182/blood-2018-99-119851.

[19] L.F Onate-Ocana, et al., Metabolic tumor volume changes assessed by interval (18) fluorodeoxyglucose positron emission tomography-computed tomography for the prediction of complete response and survival in patients with diffuse large B-cell lymphoma, Oncol Lett 16 (2018) 1411–1418, https://doi.org/10.3892/ol.2018.8817.

[20] X.U Kahle, et al., Tumour necrosis as assessed with (18)F-FDG PET is a potential prognostic marker in diffuse large B cell lymphoma independent of MYC rearrangements, Eur Radiol 29 (2019) 6018–6028, https://doi.org/10.1007/s00330-019-06178-9.

[21] B Ganeshan, et al., 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 (2017) 1012–1020, https://doi.org/10.1007/s00330-016-4470-8.

[22] B.D Cheson, et al., Recommendations for initial evaluation, staging, and response assessment of Hodgkin and non-Hodgkin lymphoma: the Lugano classification, J Clin Oncol 32 (2014) 3059–3068, https://doi.org/10.1200/JCO.2013.54.8800.

[23] A Fedorov, et al., 3D Slicer as an image computing platform for the quantitative imaging network, Magn Reson Imaging 30 (2012) 1323–1341, https://doi.org/10.1016/j.mri.2012.05.001.

[24] J.J.M van Griethuysen, et al., Computational radiomics system to decode the radiographic phenotype, Cancer Res 77 (2017) e104–e107, https://doi.org/10.1158/0008-5472.Can-17-0339.

[25] D.W Scott, et al., High-grade B-cell lymphoma with MYC and BCL2 and/or BCL6 rearrangements with diffuse large B-cell lymphoma morphology, Blood 131 (2018) 2060–2064, https://doi.org/10.1182/blood-2017-12-820605.

[26] B Chapuy, et al., Molecular subtypes of diffuse large B cell lymphoma are associated with distinct pathogenic mechanisms and outcomes, Nat Med 24 (2018) 679–690, https://doi.org/10.1038/s41591-018-0016-8.

[27] R Schmitz, et al., Genetics and pathogenesis of diffuse large B-Cell Lymphoma, New Engl J Med 378 (2018) 1396–1407, https://doi.org/10.1056/NEJMoa1801445.

[28] T Magnes, et al., Spatial Heterogeneity in Large Resected Diffuse Large B-Cell Lymphoma bulks analyzed by massively parallel sequencing of multiple Synchronous Biopsies, Cancers (Basel) 13 (2021), https://doi.org/10.3390/cancers13040650.

[29] A Younes, et al., International Working Group consensus response evaluation criteria in lymphoma (RECIL 2017), Ann Oncol 28 (2017) 1436–1447, https://doi.org/10.1093/annonc/mdx097.