

Machine-Learning Approach Using FDG-PET-Based Radiomics in the Characterization of Mediastinal Bulky Lymphomas.

Elisabetta Maria Abenavoli (✉ elisabettabenavoli@gmail.com)

AOU Careggi: Azienda Ospedaliero Universitaria Careggi <https://orcid.org/0000-0002-9370-102X>

Flavia Linguanti

AOUC: Azienda Ospedaliero Universitaria Careggi

Matteo Barbetti

INFN Firenze: Istituto Nazionale di Fisica Nucleare Sezione di Firenze

Francesco Mungai

AOUC: Azienda Ospedaliero Universitaria Careggi

Vittorio Miele

AOUC: Azienda Ospedaliero Universitaria Careggi

Luca Nassi

AOUC: Azienda Ospedaliero Universitaria Careggi

Benedetta Puccini

AOUC: Azienda Ospedaliero Universitaria Careggi

Ilaria Romano

AOUC: Azienda Ospedaliero Universitaria Careggi

Raffaella Santi

AOUC: Azienda Ospedaliero Universitaria Careggi

Alessandro Passeri

AOUC: Azienda Ospedaliero Universitaria Careggi

Roberto Sciagrà

University of Florence Department of Biochemistry: Università degli Studi di Firenze Dipartimento di Scienze Biomediche Sperimentali e Cliniche Mario Serio

Cinzia Talamonti

University of Florence Department of Experimental and Clinical Biomedical Sciences 'Mario Serio': Università degli Studi di Firenze Dipartimento di Scienze Biomediche Sperimentali e Cliniche Mario Serio

Alessandro Maria Vannucchi

University of Florence Department of Experimental and Clinical Medicine: Università degli Studi di Firenze Dipartimento di Medicina Sperimentale e Clinica

Valentina Berti

Research Article

Keywords: bulky, lymphoma, diagnosis, textural analysis, machine learning, 18 F-FDG PET/TC

Posted Date: February 9th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1330600/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Machine-Learning Approach Using FDG-PET-based Radiomics in the Characterization of Mediastinal Bulky lymphomas.

E. Abenavoli¹, F. Linguanti¹, M. Barbetti²⁻³, F. Mungai⁴, V. Miele⁴, L. Nassi⁵, B. Puccini⁵, I. Romano⁵, R. Santi⁶, A. Passeri¹, R. Sciagrà¹, C. Talamonti^{3,7}, A. M. Vannucchi⁸, V. Berti¹.

Affiliations

1. Nuclear Medicine Unit, Department of Experimental and Clinical Biomedical Sciences "Mario Serio", University of Florence, Florence, Italy
2. Department of Information Engineering, University of Florence, Florence, Italy
3. INFN, Florence Division, Florence, Italy
4. Department of Radiology, Azienda Ospedaliero Universitaria Careggi, Florence, Italy
5. Hematology Department, University of Florence and Azienda Ospedaliero Universitaria Careggi, Florence, Italy.
6. Pathology Section, Department of Health Sciences, University of Florence, Florence, Italy.
7. Medical Physics Unit, Department of Experimental and Clinical Biomedical Sciences "Mario Serio", University of Florence, Italy
8. Department of Experimental and Clinical Medicine, CRIMM, Center Research and Innovation of Myeloproliferative Neoplasms, Azienda Ospedaliero Universitaria Careggi, University of Florence, Florence, Italy.

Purpose: Mediastinal bulky involvement is common in lymphomas, particularly in classical Hodgkin lymphoma (cHL), primary mediastinal B-cell lymphoma (PMBCL) and grey zone lymphoma (GZL). Despite advanced methods of assessment, traditional imaging appears insufficient in differentiating histology. This study tested the diagnostic value of ^{18}F -FDG PET/CT volumetric and texture parameters in the histological differentiation of mediastinal bulky due to GZL, PMBCL and cHL, also using a machine-learning approach.

Methods: We retrospectively reviewed patients with mediastinal bulky disease with a histopathological diagnosis of cHL, PMBCL or GZL who underwent pre-treatment ^{18}F -FDG PET/CT. Lesions were delineated using a fully automated preselection of ^{18}F -FDG avid structures defined by a threshold $\text{SUV} \geq 2.5$. Volumetric and radiomic parameters were measured using LIFEx software both for bulky lesion (BL) and for all lesions (AL) on ^{18}F -FDG PET/CT. Analysis of selected radiomic features was performed with Machine Learning classifiers based on Logistic Regression.

Results: We reviewed 117 patients (29 PMBCL, 80 cHL, 8 GZL). The analysis showed significant differences between the 3 lymphoma groups regarding SUV_{max} , SUV_{mean} , BL/AL-MTV and BL/AL-TLG. Several PET textural features both of first order and of second order grey-level showed significant differences between the 3 groups. Finally, machine-learning classifier provided good accuracy in the discrimination between groups. Logistic Regression showed good performance, confirming true positive rate (TPR) and true negative rate (TNR) greater than 80% in the characterization of PMBCL and cHL. The multiclass classifier showed TPR greater than 70% and TNR lower 5% in the identification of PMBCL and cHL and TPR of 44% in GZL.

Conclusion: Mediastinal bulky involvement from different histologies showed diverse ^{18}F -FDG-PET radiomic features and the machine-learning approach successfully identified the different subtypes. Our results support the potential of metabolic heterogeneity as an imaging biomarker, and the use of radiomic features for early characterization of mediastinal bulky lymphoma.

Keywords: bulky, lymphoma, diagnosis, textural analysis, machine learning, ^{18}F -FDG PET/TC

Introduction

Mediastinal bulky involvement in aggressive lymphomas poses several clinical and diagnostic challenges, for reasons both related to the risks of a voluminous mass in that anatomical district and for the technical difficulties in obtaining a diagnostic sample without an aggressive surgical approach. Bulky disease is recognized as a negative prognostic factor, so its early recognition and accurate characterization is crucial in the staging process of lymphoma[1]–[4]. Several definitions and measurement methods have been suggested to define the most appropriate dimensions, but without a consensus both for Hodgkin Lymphoma (HL) and for non-Hodgkin Lymphoma (NHL)[5-8].

Another diagnostic challenge to overcome in bulky disease is biopsy. Satisfactory sampling is essential for diagnosis and classification of malignant lymphoma. Fundamental steps are identification of the best site, the easiest access, and the choice of the most appropriate technique, to avoid inconclusive results [9–11]. The surgically excised tissue biopsy is widely accepted as the gold standard for the diagnosis of lymphoma based upon the current international guidelines, however it probes only confined parts of the tumor, especially in bulky disease [1,3]. Mediastinal bulky lymphoma could be due to several lymphoma histologies, and even in the same bulky mass different histologies may coexist. Both inter- and intra-patient heterogeneity may be found, varying with the histopathological pattern of lymphoma. [2]

Finally, despite the similarities in clinical symptoms and dimensions of mediastinal bulky lymphomas, treatment strategies and clinical outcomes highly differ between the 3 most frequent histologies, which are cHL, primary mediastinal B-cell lymphoma (PMBCL) and a rare entity entitled B-cell lymphoma unclassifiable with features intermediate between diffuse large B-cell lymphoma (DLBCL) and cHL, also known as grey zone lymphoma (GZL). Because the treatment approaches are different, the pathologist must make these distinctions precisely, and the importance of differentiating histology has been highlighted. Often this differentiation is not at all simple and obvious, in particular for GZL, which differential diagnosis may be difficult even in nodal samples for a spectrum of morphologies between DLBCL and cHL with non-concordant immunohistochemical findings that may be observed within the same tumor specimen.[12]

In this scenario, morphological and functional imaging represents a fundamental diagnostic method to fully characterize bulky mediastinal lymphoma. Computed Tomography with contrast medium (ceMDCT) is the most widely used technique during staging as it allows to evaluate both superficial and deep lymph node stations with good accuracy, identifying lymph nodes with altered morphology as pathological. ¹⁸F-FDG PET/CT is also considered an indispensable tool for staging and defining the response to treatments in different lymphomas. Its role derives both from the intrinsic

characteristics of the method (glucose metabolism) capable of providing information on specific functional patterns, but also through a continuous refinement of many methodological and interpretative aspects [1,4, 6].

Currently, texture analysis is a newly developed, high throughput way to extract from images numerical information that naked eyes cannot detect and can thus quantify more image features than standard analyses [13]. Literature on functional imaging radiomics for lymphoma is still scanty. Nevertheless, published research in this field highlights two main areas of clinical interest: outcome prediction and histological differentiation from other types of cancer [13–15].

In this study, we assessed the diagnostic value of ^{18}F -FDG PET volumetric and texture parameters in the histological differentiation of mediastinal bulky lymphomas due to GZL, PMBCL and cHL. This study uses machine-learning algorithms on relevant features to predict the different histological types of the mediastinal bulky masses. The aim is to understand how PET radiomic features extracted from bulky masses may predict lymphoma histology and in the future support their histological diagnosis.

Materials and Methods

Population

From January 2010 to April 2021, a cohort of untreated patients with a histopathological diagnosis of PMBCL, cHL, or GZL was retrospectively evaluated in a monocentric study conducted by the Nuclear Medicine and Haematology divisions of Azienda Ospedaliera Universitaria Careggi (Florence, Italy). The study was approved by the Institutional ethics Review Board of the centre and all patients provided written informed consent. Eligibility criteria included ≥ 18 years, histologically proven PMBCL, cHL, and GZL. Relapse or refractory patients were excluded from the study. Demographic patient data, histology, disease stage, and presence of bulky disease were recorded from electronic patient charts and radiology information system.

All patients underwent ^{18}F -FDG PET/CT at baseline (Fig.1). Staging was assessed according to Ann Arbor classification and Lugano criteria .[6]

Image acquisition

All patients fasted for at least 4 h before examination and blood glucose levels were at least < 200 mg/dl. Images from the mid skull to pelvis were obtained on a dedicated PET/CT scanner (Philips Gemini TF 16 PET/CT), 60 min after intravenous injection of 0.1 mCi/kg of ^{18}F -FDG . Before PET scans, a low dose CT (120 kV; 50–80 mA) was acquired

to allow attenuation correction and lesions localization. PET images were reconstructed using an iterative algorithm (3D LOR RAMLA reconstruction with TOF, FOV: 576, matrix: 144×144, voxel dimension: 4×4×4 mm) and fused images of matching pairs of PET and CT images were available for review in axial, coronal, and sagittal planes and in maximum intensity projections (MIPs).

Data analysis

All available pre-treatment imaging studies, including ceMDCT and low-dose CT performed in conjunction with ¹⁸F-FDG PET imaging were reviewed. The longest diameter of the largest individual or conglomerate lymph node mass was measured in the transverse plane and coronal plane to confirm bulk disease.

PET images were independently evaluated by a nuclear medicine physician and a resident in nuclear medicine, using the open-source software platforms LifeX (Orlhac F, Nioche C, Buvat I. LIFExuser guide. LIFEx <https://www.lifexsoft.org>)[16].

Each measurement was performed with the SUV thresholding method, based on the absolute SUV value >2.5 on both bulky lesion (BL) and all lesions (AL), if present. SUV_{max} and SUV_{mean} of the BL were also evaluated, calculated in SUV_{bw} units for both variables.

To make segmentation as user independent as possible, we kept default settings suggested by the manufacturers in the user guide. The regions containing only physiological ¹⁸F-FDG uptake (brain, bladder, heart, kidney, intestine, etc.) were edited out. Furthermore, when the pathological tissue was automatically incorporated into the same volume of interest (VOI) with adjacent physiologically active tissue, the VOI was redesigned by defining the limits of the slices in which the lesion was included and redrawing it semi-automatically.

AL-MTV was defined as the sum of the metabolic volumes of every individual lesion. For each lesion, tumor lesion glycolysis was calculated as the product of the lesion volume by the SUV_{mean} within the lesion, and total lesion glycolysis was obtained by summing tumor lesion glycolysis over all lesions.

Texture analysis and Machine Learning

From PET images, we extracted the features defined as first-order and second-order, or higher-order using Local Image Features Extraction (LIFEx) software (<http://www.lifexsoft.org>) on the BL VOI.

First-order features are typically obtained from the histogram of grey-level values obtained from the considered VOIs and describe the overall variation of the signal in the VOI regardless of the relative position of the voxels. These

features are fairly invariant to geometric transforms, therefore robust with respect to image reconstruction and filtering. They include basic statistics such as mean, median, range, standard deviation, skewness, and kurtosis.

Second-order features are those that are usually referred to as *texture features*, since they consider the spatial relationship between neighbouring VOIs in an image, and thus they are capable to capture details regarding the heterogeneity of the lesions. They include parameters from Gray-level co-occurrence matrix (GLCM), neighbourhood Gray-level different matrix (NGLDM), Gray-level run length matrix (GLRLM), and Gray-level zone length matrix (GLZLM), and indices from sphericity and histogram. A detailed description of the various texture parameters evaluated can be found at <http://www.lifexsoft.org>. Conventional and advanced metabolic tumor parameters as well as radiomics features assessed are summarized in **Table 1**.

Radiomic features from PET/CT images were fed into Machine Learning models in order to build a histological representation of mediastinal lymphomas, able to help during the diagnostic phase [17–19]. The use of Machine Learning techniques assumes that the training dataset is statistically significant, not only in terms of instance multiplicity but also in terms of the discriminant power brought by each instance [20–22]. Thus, before training any models, an intense data cleaning over texture features has been performed, removing all the features with strong correlations ($|PCC| > 0.75$) and the ones without any discriminant power.

Lymphoma type identification pushes the Machine Learning technique also due to the separation of non-completely disjoint histological classes. This histological-based limitation prevents to directly perform multiclass studies, requiring more complex solution for classification. Binary classification between PMBCL and cHL was carried out, then the resulting classifier was enhanced with multiclass capabilities thanks to the definition of decision regions embedding the mixed characteristics of GZL.

Statistical analysis

Statistical analysis was performed with the SPSS software (SPSS25) by means of ANOVA to evaluate the differences of the means between the measured variables. Quantitative data are presented as mean \pm standard deviation (SD). A p value <0.05 was considered significant.

RESULTS

Patients

One-hundred thirty-one patients with histologically proven PMBCL, cHL or GZL were included in the study. Fourteen patients were excluded from the study because of missing ^{18}F -FDG PET/CT at baseline. Seven of the analyzed patients (all cHL) were excluded from the clinical evaluation because they were treated outside of our Institution but were included in the texture assessment for availability of staging ^{18}F -FDG PET/CT. Patient characteristics are summarized in Table 2.

Twenty-nine patients were PMBCL (24.8%), 80 (68.4%) were cHL and 8 (6.8%) were GZL. The median age for PMBCL was 40 years (range, 21-59), for cHL was 33 years (range, 18-74) in the 73 available patients and for GZL was 47 years (range, 18-60).

In this study we considered as tumor bulk all the mediastinal masses with a diameter ≥ 5 cm. Median largest mediastinal diameter measured for PMBCL was 11.0 cm (range, 6.5-17), 8.9 cm (range, 5-20) in cHL patients and 13.5 cm (range, 5-16) measured in GZL patients. Elevated lactate dehydrogenase (LDH) was observed especially for PMBCL, with median values of 355 UI/l (range, 134-757), evaluated in 27 patients. Median erythrocyte sedimentation rate (ERS) was 37 mm/h (range, 6-106), 31 mm/h (range, 4-120) and 34.5 mm/h (range, 20-62) for PMBCL, cHL and GZL, respectively.

Radiomic features

Several conventional and advanced metabolic parameters, as well as texture features, were significantly different among the three groups. SUV_{max} and TLG were significantly different among groups, with highest values in PMBCL, intermediate in GZL and lowest in cHL (Fig.2). PMBCL and GZL had similar MTV values, but significantly higher than cHL.

Among first order variables, histogram skewness and kurtosis were significantly different among groups ($p < 0.001$). PMBCL showed significantly lower (negative) values as compared to the other groups, indicating asymmetrical histograms with left-sided skewness. On the contrary, cHL showed significantly higher positive values than the other groups, indicating asymmetrical histograms with right-sided skewness.

Regarding kurtosis, cHL showed significantly higher kurtosis values, indicating a value distribution with heavier tail, or a greater number of outliers, as compared to PMBCL, which showed significantly lower kurtosis values.

Among the features extracted from NGLDM, only Contrast resulted significantly different, with PMBCL showing significantly higher contrast than GZ and cHL, and GZL significantly higher than cHL.

In GLZLM, which provided information on the size of homogeneous zones for each Gray-level in 3 dimensions, several texture features resulted significantly different among groups (Fig.3). SZE, HGZE, SZHE and ZLNU parameters showed significantly lower values in cHL, intermediate in GZL and higher values in PMBCL. Conversely, LZE, LGZE and SZLGE parameters showed decreasing values between groups, with significantly higher values in cHL, as compared to both GZL and PMBCL.

GLCM features, which depicts the occurrence rate of the brightness levels on the image at fixed distances and orientations, resulted significantly different among groups ($p < 0.001$); variance, entropy and dissimilarity showed lower values in cHL, higher values in PMBCL and intermediate values in GZL. On the other hand, GLCM homogeneity resulted significantly lower in cHL, intermediate in GZL and higher in PMBCL.

Among GLRLM features, LRE, LRLGE, SRLGE, HGRE and RP showed significantly lower values in cHL, intermediate values in GZL and higher values in PMBCL. Instead, HGRE, SRHGE and SRE registered higher values in cHL, intermediate in GZL and lower values in PMBCL. Regarding RLNU, PMBCL and GZL showed similar values, both significantly higher than cHL.

Machine Learning Analysis

The Machine Learning analysis started with the reduction of the size of the dataset for training and testing models to 101 (28 PMBCL, 65 cHL and 8 GZL), because of missing data for some features of the study.

The first step of lymphoma type inference was the binary classification PMBCL-cHL. A stratified splitting in training and test sets (80/20) was performed, resulting in 73 instances (21 PMBCL and 52 cHL) to train models. Due to the small number of instances for training, we used a simple classification technique like Logistic Regression, implemented in Python through the Scikit-Learn package [23]. The classifier was trained on 17 parameters including conventional and advanced metabolic indices, together with texture and other features listed in Table 1. The five most important features for lymphomas discrimination are reported in Table 3, together with the list of parameters dropped due to their correlations [24].

Once trained the model, the decision boundary for predictions was set to maximize the PMBCL identification (TP) within the training set, while ensuring a tolerable misidentification for cHL (FP). Logistic Regression with custom

decision boundary showed good performance on the testing set as well, confirming a TPR greater than 80% as reported on the confusion matrix of Figure 4A.

GZL predictions have been added to the binary classifier capabilities with the construction of three decision regions. Since its mixed characteristics, the GZ region was built around the binary decision boundary with an asymmetric interval (Fig. 4B).

As reported in the confusion matrix, the so-built multiclass classifier showed TPR greater than 70% in the identification of PMBCL and cHL. GZL was identified with TPR of 44% (Fig.4C).

Discussion

In this study we demonstrated the diagnostic value of ^{18}F -FDG PET volumetric and texture parameters in the histological differentiation of mediastinal bulky lymphomas due to GZL, PMBCL and cHL. Through appropriate features, the machine learning algorithms predicted the different histological types of the mediastinal bulky masses with acceptable sensitivity in the identification of PMBCL and cHL.

Among conventional quantitative variables, SUV measurements resulted significantly different among the three lymphoma histologies, with lower values in cHL, intermediate in GZL and higher in PMBCL. In literature, SUVs have been the first and most investigated measures to provide objective quantification of tumor biological activity. Several studies in lymphoma patients correlated ^{18}F -FDG activity with tumor proliferative ability, showing that high-proliferative lymphomas, such as PMBCL, often shows very high metabolic activity [25–27]. Our results are consistent with these findings, confirming that differences in proliferative activity may be a possible explanation for the different ^{18}F -FDG uptake in various histotypes. Moreover, both SUV_{max} and SUV_{mean} values were significantly lower in cHL compared with PMBCL and GZL, consistently with its peculiar neoplastic tissue architecture, composed of few and scattered neoplastic cells called Hodgkin and Reed-Sternberg (HRS) cells, accounting for less than 1% of the total cell count, surrounded by an overwhelming population of non-neoplastic mononuclear bystander cells with a high metabolism [28, 29]. Such histopathological features may explain the differences in metabolic activity of cHL as compared to PMBCL. Finally, we found intermediate behaviour in GZL group, again consistent with its biological characteristics. Indeed, GZL could show a wide range of cellular patterns, with coexistence of cells resembling HRS cells, cells with centroblastic or immunoblastic cytology and marked nuclear pleomorphism, and cells with more monomorphic cytology, resembling PMBCL [30].

Beyond SUV values, there were significant differences among lymphoma types also regarding volume-based variables, such as MTV and TLG. MTV represents one important prognostic factor in lymphoma, beside stage and the other clinical factors such as LDH, Performance Status, age and presence of extranodal disease. Not surprisingly, most studies have demonstrated that MTV was an adverse prognostic factor, regardless of lymphoma histology and choice of quantification method [8,31–33].

Interestingly, our data are in line with the concept that volumetric parameters have greater values in aggressive histotypes, with consequent worse prognosis. Indeed, in our study PMBCL and GZL showed higher values of both MTV and TLG than cHL and interestingly PMBCL significantly differ from GZL in TLG.

In addition, one of the most interesting findings our study is the significant difference between lymphoma types regarding several texture features extracted from ^{18}F -FDG PET images. Metabolic heterogeneity is a remarkable feature of malignancy, and texture analysis provides tools for quantifying tumor heterogeneity [34]. Those features measuring increasing heterogeneity within tumors may be associated with differences in regional tumor cellularity, proliferation, hypoxia, angiogenesis, and necrosis, as well as genomic variety, all factors that have been independently associated with more aggressive behaviour, poorer response to treatment and worse prognosis [35–37].

We chose to focus on patients presenting with mediastinal bulky lymphoma since it is well-established that the complementary information provided by textural analysis strongly depends on tumor size. Although the information increases substantially with larger volumes, however, the level of correlation between the variables tends to decrease substantially [30, 38].

In our study, the high co-occurrences features, analysing inter-relationships between pairs of voxels and thus characterizing local non-uniformities, were able to significantly differentiate the three lymphoma groups. Orhac et al. reported that $\text{LZE}_{\text{GLZLM}}$ exhibited higher values in visually homogeneous lesions than in heterogeneous lesions [39]. They identified two different sets of texture indices able to reflect the different characteristics of uptake heterogeneity. The first one included homogeneity and entropy derived from GLCM, and run emphasis features derived from GLRLM, all of which were sensitive to the presence of uptake heterogeneity. The second set comprised gray-level zones for GLZLM, which were mostly sensitive to the average uptake rather than to the local heterogeneity. So homogeneous lesions like cHL had a higher value of $\text{homogeneity}_{\text{GLCM}}$ and $\text{LGZE}_{\text{GLZLM}}$ and a lower value of $\text{entropy}_{\text{GLCM}}$ and $\text{HGZE}_{\text{GLZLM}}$ than heterogeneous lesions as PMBCL and GZL. As expected, entropy, which reflects disorder, varies in an opposite direction to homogeneity. Another important aspect is how SUV measurements and entropy are strongly correlated and this is also reflected in the trend of our results, with higher values in PMBCL and lower in cHL [40, 41]. The more chaotic the image is the higher the entropy [13, 42]. The other features evaluated in

this study highlight tumor heterogeneity, depending on the type of matrix used and the kind of feature computed on this matrix. Consequently, whereas a single feature cannot be directly linked to a specific biologic process, one could assume that a combination of textural parameters may be closely related to underlying physiologic processes, such as vascularization, neoangiogenesis, perfusion, tumor aggressiveness, or hypoxia [43, 44].

Finally referring to the GLZM matrix, the feature ZLNU, which measures the variability of the intensity values of the gray level in the image, indicated greater homogeneity in cHL than in PMBCL, showing on the contrary very high values.

Literature studying the possibility of pathological predictions of histological subtypes based on radiomic features from ¹⁸F-FDG PET/CT images have produced conflicting results. Previous studies excluded the possibility of creating a close correlation between histology and radiomics, even if texture indices may provide some useful information in the spatial organization of tumor cells[45]. More recently, the application of machine learning methods to radiomics allowed to produce a classification modelling to predict the different subtypes of malignant lymphomas[17,46]. Our data are consistent with these latter results, demonstrating that machine learning analyses can create a classification model with good discriminant power.

Of all available indices, we observed that all top five features (SUV_{mean} , MTV, SRE_{GLRLM} , $COARSENESS_{NGDLM}$, $ZLNU_{GLZLM}$) obtained with our best performing model (Logistic regression) are mostly activity dependent, suggesting that metabolic and average differences regardless of spatial relationships are essential for the differentiation of tumor lesions. Nevertheless, Machine Learning can fulfill good performances only through the synergism between metabolic activity parameters and high order features, confirming the complex discrimination of biological processes in pathological tissue. The other selected features are reasonably considered relevant in discrimination, but with less direct influence in the algorithms' performances.

The so-built multiclass classifier showed promising performance in the separation of the three lymphoma types, paying with an unavoidable drop in performance with respect to the binary classification. The model assumption that GZL has intermediate characteristics between cHL and PMBCL derives from its biological traits and is fully reflected by the observed distribution radiomic variables (Fig. 4). Therefore, despite the presence of histological subtypes that are not completely disjointed, Logistic Regression was able to obtain very good classification results: 1) very high TPR value in the identification of PMBCL, which represents the histotype with the worst prognosis, and therefore the most significant one to be identified by haematologists; 2) just under 5% of PMBCL are erroneously classified as cHL; 3) the discrimination values of cHL are slightly lower than PMBCL, but nevertheless valid (TPR=73%).

Our radiomic data, in conjunction with machine learning analyses, demonstrate that textural features derived from PET imaging correlate with histologic diagnosis in patients with mediastinal bulky lymphoma and thus may support anatomopathological analyses. In this context, ^{18}F -FDG PET could aim to be used in several additional ways over the standard clinical use. On one hand, it could represent the best method to guide biopsy in a voluminous bulky mass (where and how many samples?), eventually identifying the metabolic patterns possibly corresponding to specific histologies. On the other hand, in the future ^{18}F -FDG PET radiomics could help clarifying possible discordances between the histopathological diagnosis and clinical characteristics of the patient. In this last case, radiomic features of the bulky mass could represent an additional variable that could help the clinicians and the pathologists to define the diagnosis or guide the decision to perform a second biopsy in a more suspect area of the bulky lesion.

The present study had some limitations. First, this was a retrospective study. Secondly, PMBCL and GZL cohorts are limited, especially as compared to cHL group; however, such proportions respect the incidence of these diseases. Thus, it is important to validate these results with an enlarged sample size or a prospective trial in the future. Finally, the relationship between texture parameters and histopathological findings requires further study, to find out each tumor biological dependency of each textural feature.

CONCLUSION

Different subtypes of bulky mediastinal lymphoma, namely cHL, GZL and PMBCL, demonstrated different ^{18}F -FDG PET characteristics and metabolic heterogeneity examined by radiomics and texture analysis. Machine learning could be successfully combined with radiomics, providing good discriminative sensitivity, especially for cHL, and PMBCL. In the era of precision medicine and personalized treatment, this preliminary study supports the potential of metabolic texture analyses as future imaging biomarker, with growing role in the clinical diagnosis.

REFERENCES

- [1] S. H. Swerdlow *et al.*, "The 2016 revision of the World Health Organization classification of lymphoid neoplasms," *Blood*, vol. 127, no. 20. 2016. doi: 10.1182/blood-2016-01-643569.
- [2] C. M. Schürch, B. Federmann, L. Quintanilla-Martinez, and F. Fend, "Tumor Heterogeneity in Lymphomas: A Different Breed," *Pathobiology*, vol. 85, no. 1–2, 2018, doi: 10.1159/000475530.
- [3] G. Stanta and S. Bonin, "Overview on clinical relevance of intra-tumor heterogeneity," *Frontiers in Medicine*, vol. 5, no. APR. 2018. doi: 10.3389/fmed.2018.00085.

- [4] B. D. Cheson, "PET/CT in Lymphoma: Current Overview and Future Directions," *Seminars in Nuclear Medicine*, vol. 48, no. 1. 2018. doi: 10.1053/j.semnuclmed.2017.09.007.
- [5] D. A. Eichenauer *et al.*, "Hodgkin lymphoma: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up," *Annals of Oncology*, vol. 29, 2018, doi: 10.1093/annonc/mdy080.
- [6] B. D. Cheson *et al.*, "Recommendations for initial evaluation, staging, and response assessment of hodgkin and non-hodgkin lymphoma: The lugano classification," *Journal of Clinical Oncology*, vol. 32, no. 27. 2014. doi: 10.1200/JCO.2013.54.8800.
- [7] M. Pfreundschuh *et al.*, "Prognostic significance of maximum tumour (bulk) diameter in young patients with good-prognosis diffuse large-B-cell lymphoma treated with CHOP-like chemotherapy with or without rituximab: an exploratory analysis of the MabThera International Trial Group (MInT) study," *The Lancet Oncology*, vol. 9, no. 5, 2008, doi: 10.1016/S1470-2045(08)70078-0.
- [8] U. Tateishi *et al.*, "Prognostic significance of metabolic tumor burden by positron emission tomography/computed tomography in patients with relapsed/refractory diffuse large B-cell lymphoma," *Cancer Science*, vol. 106, no. 2, 2015, doi: 10.1111/cas.12588.
- [9] S. Gupta *et al.*, "Imaging-guided percutaneous biopsy of mediastinal lesions: Different approaches and anatomic considerations," *Radiographics*, vol. 25, no. 3. 2005. doi: 10.1148/rg.253045030.
- [10] Z. Wang and C. Jiang, "Endoscopic ultrasound in the diagnosis of mediastinal diseases," *Open Medicine (Poland)*, vol. 10, no. 1. 2015. doi: 10.1515/med-2015-0095.
- [11] L. Li *et al.*, "Value of CT-guided core-needle biopsy in diagnosis and classification of malignant lymphomas using automated biopsy gun," *World Journal of Gastroenterology*, vol. 11, no. 31, 2005, doi: 10.3748/wjg.v11.i31.4843.
- [12] J. Bruneau and T. J. Molina, "WHO Classification of Tumors of Hematopoietic and Lymphoid Tissues," 2020. doi: 10.1007/978-3-319-95309-0_3817.
- [13] Y. W. Sun *et al.*, "Differentiating gastric cancer and gastric lymphoma using texture analysis (TA) of positron emission tomography (PET)," *Chinese medical journal*, vol. 134, no. 4, 2020, doi: 10.1097/CM9.0000000000001206.
- [14] S. Zhu *et al.*, "differential diagnostic ability of 18f-fdg PeT/cT radiomics features between renal cell carcinoma and renal lymphoma," *Quarterly Journal of Nuclear Medicine and Molecular Imaging*, vol. 65, no. 1, 2021, doi: 10.23736/S1824-4785.19.03137-6.
- [15] B. Feng *et al.*, "A Heterogeneity Radiomic Nomogram for Preoperative Differentiation of Primary Gastric Lymphoma From Borrmann Type IV Gastric Cancer," *Journal of computer assisted tomography*, vol. 45, no. 2, 2021, doi: 10.1097/RCT.0000000000001117.
- [16] F. Orhac, C. Nioche, I. Klyuzhin, A. Rahmim, and I. Buvat, "Radiomics in PET Imaging:: A Practical Guide for Newcomers," *PET Clinics*, vol. 16, no. 4. 2021. doi: 10.1016/j.cpet.2021.06.007.
- [17] F. M. de Jesus *et al.*, "Machine learning in the differentiation of follicular lymphoma from diffuse large B-cell lymphoma with radiomic [18F]FDG PET/CT features," *European Journal of Nuclear Medicine and Molecular Imaging*, Dec. 2021, doi: 10.1007/s00259-021-05626-3.
- [18] J. G. Richens, C. M. Lee, and S. Johri, "Improving the accuracy of medical diagnosis with causal machine learning," *Nature Communications*, vol. 11, no. 1, 2020, doi: 10.1038/s41467-020-17419-7.
- [19] E. J. Topol, "High-performance medicine: the convergence of human and artificial intelligence," *Nature Medicine*, vol. 25, no. 1. 2019. doi: 10.1038/s41591-018-0300-7.

- [20] K. R. Foster, R. Koprowski, and J. D. Skufca, "Machine learning, medical diagnosis, and biomedical engineering research - commentary," *BioMedical Engineering Online*, vol. 13, no. 1. 2014. doi: 10.1186/1475-925X-13-94.
- [21] M. J. J. Douglass, "Book Review: Hands-on Machine Learning with Scikit-Learn, Keras, and Tensorflow, 2nd edition by Aurélien Géron," *Physical and Engineering Sciences in Medicine*, vol. 43, no. 3, 2020, doi: 10.1007/s13246-020-00913-z.
- [22] B. Koçak, E. Ş. Durmaz, E. Ateş, and Ö. Kılıçkesmez, "Radiomics with artificial intelligence: A practical guide for beginners," *Diagnostic and Interventional Radiology*, vol. 25, no. 6. 2019. doi: 10.5152/dir.2019.19321.
- [23] F. Pedregosa *et al.*, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, 2011.
- [24] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, "Gene selection for cancer classification using support vector machines," *Machine Learning*, vol. 46, no. 1–3, 2002, doi: 10.1023/A:1012487302797.
- [25] H. Minn, H. Joensuu, A. Ahonen, and P. Klemi, "Fluorodeoxyglucose imaging: A method to assess the proliferative activity of human cancer in vivo. Comparison with DNA flow cytometry in head and neck tumors," *Cancer*, vol. 61, no. 9, 1988, doi: 10.1002/1097-0142(19880501)61:9<1776::AID-CNCR2820610909>3.0.CO;2-7.
- [26] K. Higashi *et al.*, "FDG PET measurement of the proliferative potential of non-small cell lung cancer," *Journal of Nuclear Medicine*, vol. 41, no. 1, 2000.
- [27] Y. Shou, J. Lu, T. Chen, D. Ma, and L. Tong, "Correlation of fluorodeoxyglucose uptake and tumor-proliferating antigen Ki-67 in lymphomas," *Journal of Cancer Research and Therapeutics*, vol. 8, no. 1, 2012, doi: 10.4103/0973-1482.95182.
- [28] M. Riester, Q. Xu, A. Moreira, J. Zheng, F. Michor, and R. J. Downey, "The Warburg effect: Persistence of stem-cell metabolism in cancers as a failure of differentiation," *Annals of Oncology*, vol. 29, no. 1, 2018, doi: 10.1093/annonc/mdx645.
- [29] T. Álvaro *et al.*, "Tumor-infiltrated immune response correlates with alterations in the apoptotic and cell cycle pathways in Hodgkin and Reed-Sternberg cells," *Clinical Cancer Research*, vol. 14, no. 3, 2008, doi: 10.1158/1078-0432.CCR-07-1246.
- [30] M. Hatt *et al.*, "18F-FDG PET uptake characterization through texture analysis: Investigating the complementary nature of heterogeneity and functional tumor volume in a multi-cancer site patient cohort," *Journal of Nuclear Medicine*, vol. 56, no. 1, 2015, doi: 10.2967/jnumed.114.144055.
- [31] M. Sasanelli *et al.*, "Pretherapy metabolic tumour volume is an independent predictor of outcome in patients with diffuse large B-cell lymphoma," *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 41, no. 11, 2014, doi: 10.1007/s00259-014-2822-7.
- [32] K. Nakagawa, S. Takahashi, M. Endo, Y. Ohde, H. Kurihara, and T. Terauchi, "Can 18F-FDG PET predict the grade of malignancy in thymic epithelial tumors? An evaluation of only resected tumors," *Cancer Management and Research*, vol. 9, pp. 761–768, Dec. 2017, doi: 10.2147/CMAR.S146522.
- [33] M. K. Song *et al.*, "Clinical significance of metabolic tumor volume by PET/CT in stages II and III of diffuse large B cell lymphoma without extranodal site involvement," *Annals of Hematology*, vol. 91, no. 5, 2012, doi: 10.1007/s00277-011-1357-2.

- [34] S. Ha *et al.*, "Autoclustering of Non-small Cell Lung Carcinoma Subtypes on 18F-FDG PET Using Texture Analysis: A Preliminary Result," *Nuclear Medicine and Molecular Imaging*, vol. 48, no. 4, 2014, doi: 10.1007/s13139-014-0283-3.
- [35] S. Chicklore, V. Goh, M. Siddique, A. Roy, P. K. Marsden, and G. J. R. Cook, "Quantifying tumour heterogeneity in 18F-FDG PET/CT imaging by texture analysis," *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 40, no. 1. 2013. doi: 10.1007/s00259-012-2247-0.
- [36] A. Jackson, J. P. B. O'Connor, G. J. M. Parker, and G. C. Jayson, "Imaging tumor vascular heterogeneity and angiogenesis using dynamic contrast-enhanced magnetic resonance imaging," *Clinical Cancer Research*, vol. 13, no. 12. 2007. doi: 10.1158/1078-0432.CCR-07-0238.
- [37] M. Diehn *et al.*, "Identification of noninvasive imaging surrogates for brain tumor gene-expression modules," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, no. 13, 2008, doi: 10.1073/pnas.0801279105.
- [38] F. ben Bouallègue, Y. al Tabaa, M. Kafrouni, G. Cartron, F. Vauchot, and D. Mariano-Goulart, "Association between textural and morphological tumor indices on baseline PET-CT and early metabolic response on interim PET-CT in bulky malignant lymphomas," *Medical Physics*, vol. 44, no. 9, 2017, doi: 10.1002/mp.12349.
- [39] F. Orhac, C. Nioche, M. Soussan, and I. Buvat, "Understanding changes in tumor texture indices in PET: A comparison between visual assessment and index values in simulated and patient data," *Journal of Nuclear Medicine*, vol. 58, no. 3, 2017, doi: 10.2967/jnumed.116.181859.
- [40] a Gebejes, E. M. Master, and a Samples, "Texture Characterization based on Grey-Level Co-occurrence Matrix," *Conference of Informatics and Management Sciences*, 2013.
- [41] S. Tan *et al.*, "Spatial-temporal [18F]FDG-PET features for predicting pathologic response of esophageal cancer to neoadjuvant chemoradiation therapy," *International Journal of Radiation Oncology Biology Physics*, vol. 85, no. 5, 2013, doi: 10.1016/j.ijrobp.2012.10.017.
- [42] F. Tixier *et al.*, "Intratumor heterogeneity characterized by textural features on baseline 18F-FDG PET images predicts response to concomitant radiochemotherapy in esophageal cancer," *Journal of Nuclear Medicine*, vol. 52, no. 3, 2011, doi: 10.2967/jnumed.110.082404.
- [43] M. Kunkel *et al.*, "Overexpression of Glut-1 and increased glucose metabolism in tumors are associated with a poor prognosis in patients with oral squamous cell carcinoma," *Cancer*, vol. 97, no. 4, 2003, doi: 10.1002/cncr.11159.
- [44] J. Kazmierska *et al.*, "Assessment of tumour hypoxia, proliferation and glucose metabolism in head and neck cancer before and during treatment," *The British journal of radiology*, vol. 93, no. 1106, 2020, doi: 10.1259/bjr.20180781.
- [45] F. Orhac, B. Thézé, M. Soussan, R. Boisgard, and I. Buvat, "Multiscale Texture Analysis: From 18F-FDG PET Images to Histologic Images," *Journal of Nuclear Medicine*, vol. 57, no. 11, 2016, doi: 10.2967/jnumed.116.173708.
- [46] M. Lippi *et al.*, "Texture analysis and multiple-instance learning for the classification of malignant lymphomas," *Computer Methods and Programs in Biomedicine*, vol. 185, 2020, doi: 10.1016/j.cmpb.2019.105153.

Figure Legend

Figure 1: Baseline ^{18}F -FDG-PET/CT in patient with a) cHL b) GZL c) PMBCL

Figure 2: Histograms representing mean SUV_{max} and SUV_{mean} (a), MTV and MTV_{TOT} (b), TLG and TLG_{tot} (c) in the subsets of cHL, GZL and PMBCL

Figure 3: Histograms representing mean LRE and SRE derived from GLRLM (a), homogeneity and entropy derived from GLCM (b), LGZE, HGZE, SZHGE and ZLNU derived from GLZLM (c) in the subsets of cHL, GZL and PMBCL

Figure 4: a) Normalized confusion matrix reporting the performance on the testing set (6 PMBCL and 13 cHL) of the Logistic Regression with custom boundary decision. The values reported on the diagonal are the TNR and TPR respectively (top-down), while the off-diagonal values are the FPR and FNR respectively (top-down).b) Training instances, with the addition of the GZL ones, are reported against the score achieved by the binary classifier. As expected on a histological basis, the GZL instances result in the middle of PMBCL-cHL classification. Opening an asymmetrical band around the binary decision boundary, it is possible to build three decision regions for multiclass predictions. c) Normalized confusion matrix reporting the performance on the training set (21 PMBCL, 52 cHL and 8 GZL) of the Logistic Regression with two custom decision boundaries. The values reported on the diagonal are the sensitivity with respect to the three labels: cHL, GZL and PMBCL respectively.

Tables

Table1. Parameters evaluated included conventional and advanced metabolic indices, and texture and other features

Index Matrix Parameter	Matrix	Parameter
Conventional indices		SUVmean, SUVmax, SUVpeak, SUVStd
Volumetric indices		MTV, TLG
Texture features: first order	Histogram	Skewness, kurtosis, entropy, energy
Texture features: second order	GLCM	homogeneity, energy, contrast, correlation, entropy, dissimilarity
	NGLDM	coarseness, contrast, busyness
	GLRLM	SRE/LRE, LGRE/HGRE, SRLGE/SRHGE, LRLGE/LRHGE, GLNUr/RLNU, RP
	GLZLM	SZE, LZE, LGZE, HGZE, SZLGE, SZHGE, LZLGE, LZHGE, GLNUz, ZLNU, ZP

GLCM gray-level co-occurrence matrix, NGLDM neighborhood gray-level difference matrix, GLRLM gray-level run length matrix, SRE/LRE short/long-run emphasis, LGRE/HGRE low/high gray-level run emphasis, SRLGE/SRHGE short run low/high gray-level emphasis, LRLGE/ LRHGE long run low/high gray-level emphasis, GLNUr/RLNU gray-level non-uniformity for run/run length non-uniformity, RP run percentage, GLZLM gray-level zone length matrix, SZE/LZE short/long-zone emphasis, LGZE/HGZE low/high gray-level zone emphasis, SZLGE/SZHGE short-zone low/high gray-level emphasis, LZLGE/LZHGE long-zone low/ high gray-level emphasis, GLNUz/ZLNU gray-level nonuniformity for zone or zone length nonuniformity, ZP zone percentage

Table 2. summary of the study population

Characteristics	PMBCL, n (%)	cHL, n (%)	GZL, n (%)
Patients	29 (24.8%)	80 (68.4%)	8 (6.8%)
Male sex	10 (34.5%)	35 (37%)	5 (62.5%)
Median age y (range)	40 (21-59)	33 (17.5-74)	47 (16-60).
Ann Arbor Stage			
I	6 (21%)	1 (2%)	0 (0%)
II	12 (58%)	48 (66%)	7 (87.5%)
III	2 (7%)	12 (16%)	1 (12.5)
IV	4 (14%)	12 /16%)	0 (0%)
B symptoms	8 (28%)	29 (40%)	4 (50%)
LDH UI/l (range)	355 (134-757)	207 (135-630)	255.5 (206-335)
ERS mm/h (range)	37 (6-106)	321(4-120)	34.5 (20-62)
Median bulky diameter cm (range)	11.0 (6.5-17)	8.9 (5-20)	13.5 (5-20)

Abbreviations: LDH, lactate dehydrogenase; ESR, Erythrocyte Sedimentation Rate; PMBCL, primary mediastinal B-cell lymphoma; cHL, classical Hodgkin lymphoma; GZL, grey zone lymphoma

Statements and Declarations

Conflict of interest. The authors declare that no funds, grants, or other support were received during the preparation of this manuscript

Competing Interests. The authors have no relevant financial interests to disclose.

Ethical approval. All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards (31/11/2021 CEAVC 21101)

Author contribution E. Abenavoli, F. Linguanti, M. Barbetti, F. Mungai, L. Nassi and I. Romano conception and design, acquisition of data, literature search, analysis/interpretation of data and writing of the manuscript. V. Berti, A.M. Vannucchi, C. Talamonti, R. Sciagrà, B. Puccini and V. Miele conception and design, analysis/interpretation of data, supervision, and writing of the manuscript. A. Passeri and R. Santi conception and design, acquisition of data, interpretation of data, writing of the manuscript.

Informed consent. Informed consent was obtained from the participant included in the study

Consent to publish. The authors affirm that human research participants provided informed consent for publication of the images in Figure(s) 1a, 1b and 1c.

Data and materials availability. Anonymized data for our analyses presented in this report are available upon request from the corresponding authors.

Figures

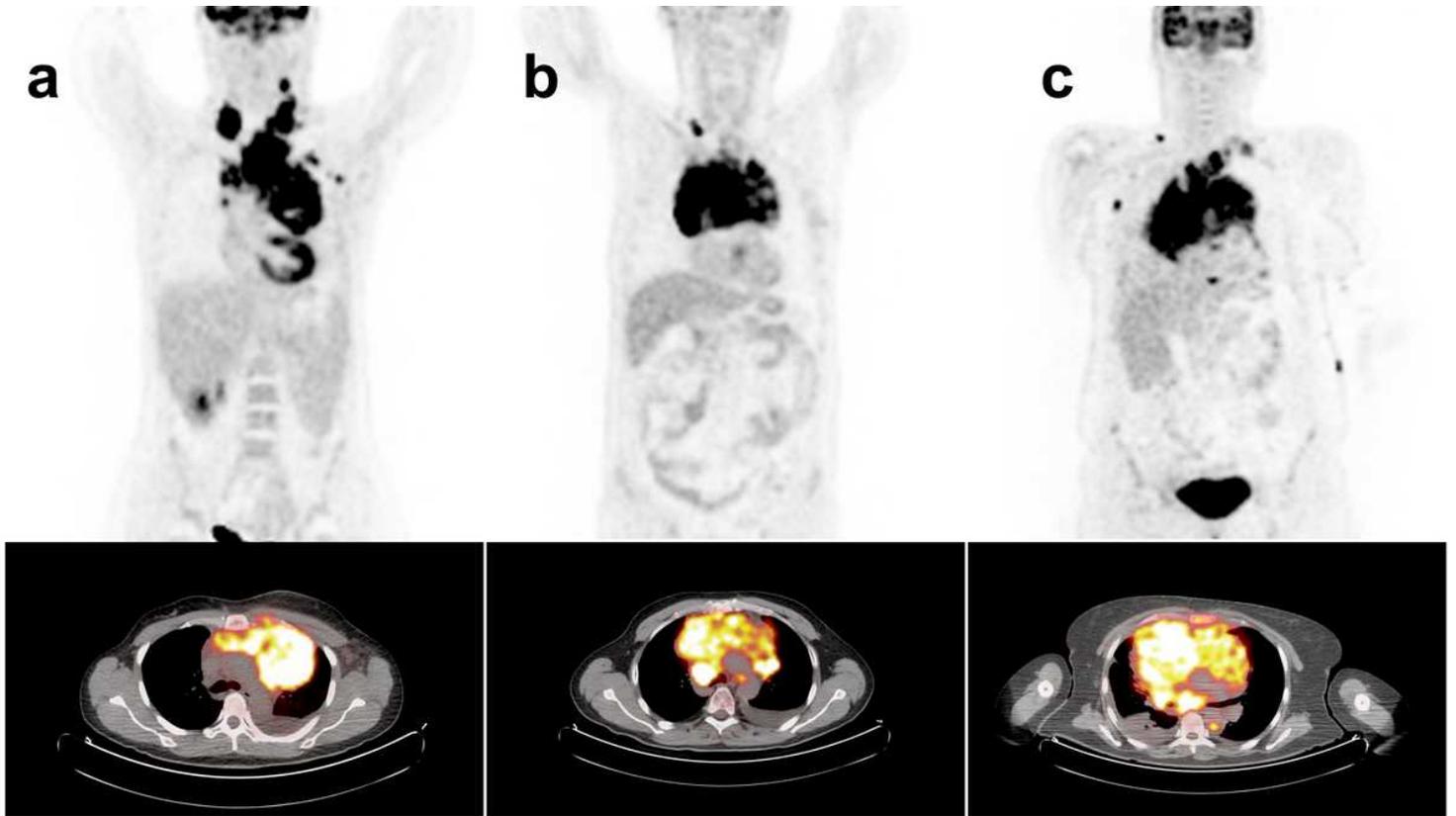


Figure 1

Baseline ^{18}F -FDG-PET/CT in patient with a) cHL b) GZL c) PMBCL

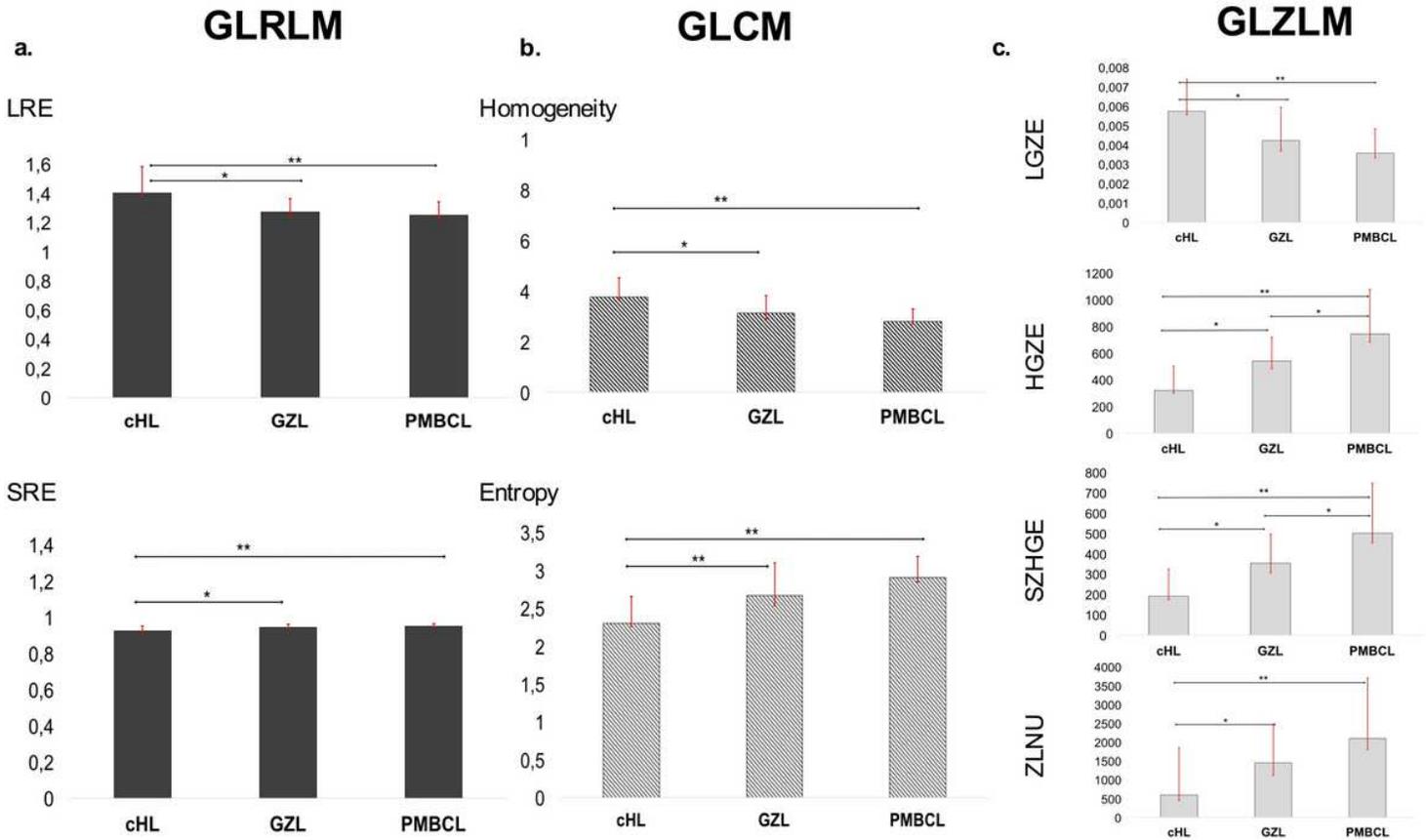


Figure 2

Histograms representing mean SUV_{max} and SUV_{mean} (a), MTV and MTV_{TOT} (b), TLG and TLG_{tot} (c) in the subsets of cHL, GZL and PMBCL

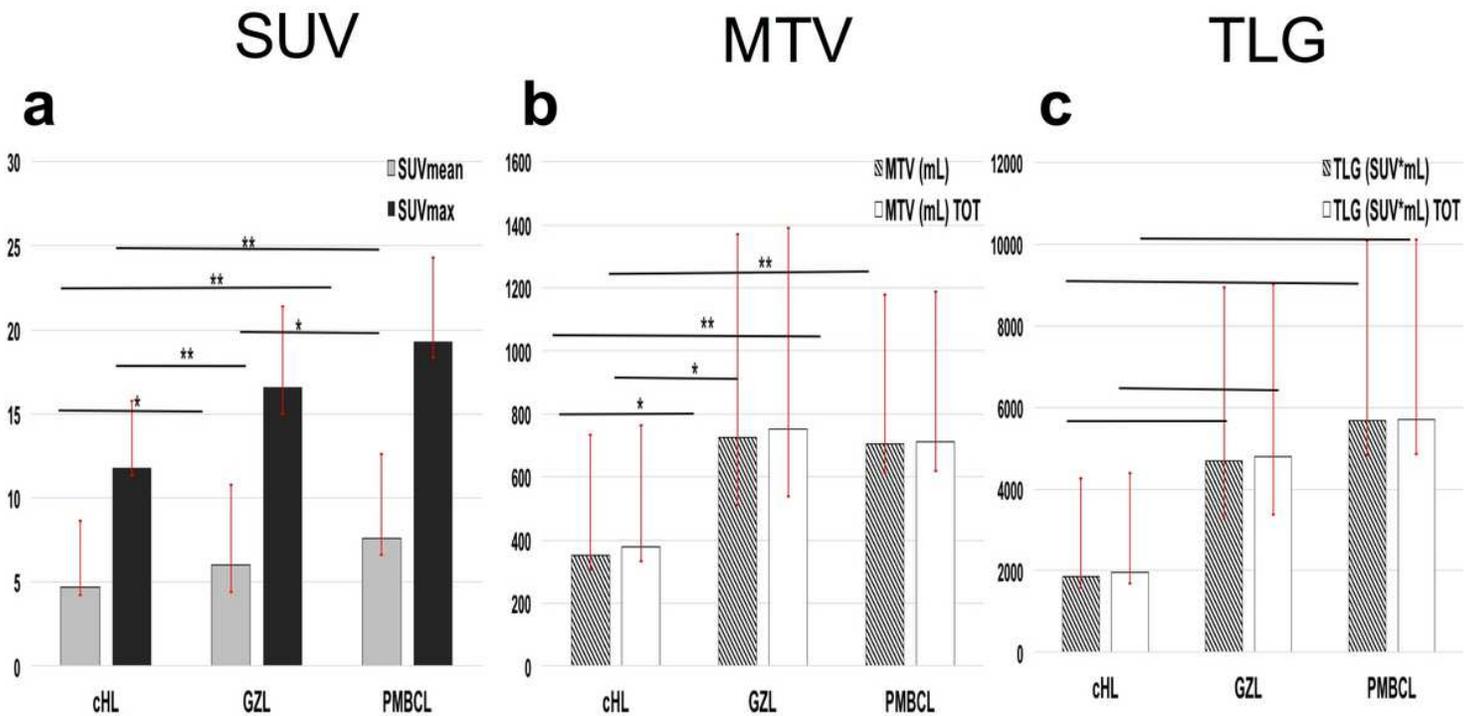


Figure 3

Histograms representing mean LRE and SRE derived from GLRLM (a), homogeneity and entropy derived from GLCM (b), LGZE, HGZE, SZHGE and ZLNU derived from GLZLM (c) in the subsets of cHL, GZL and PMBCL

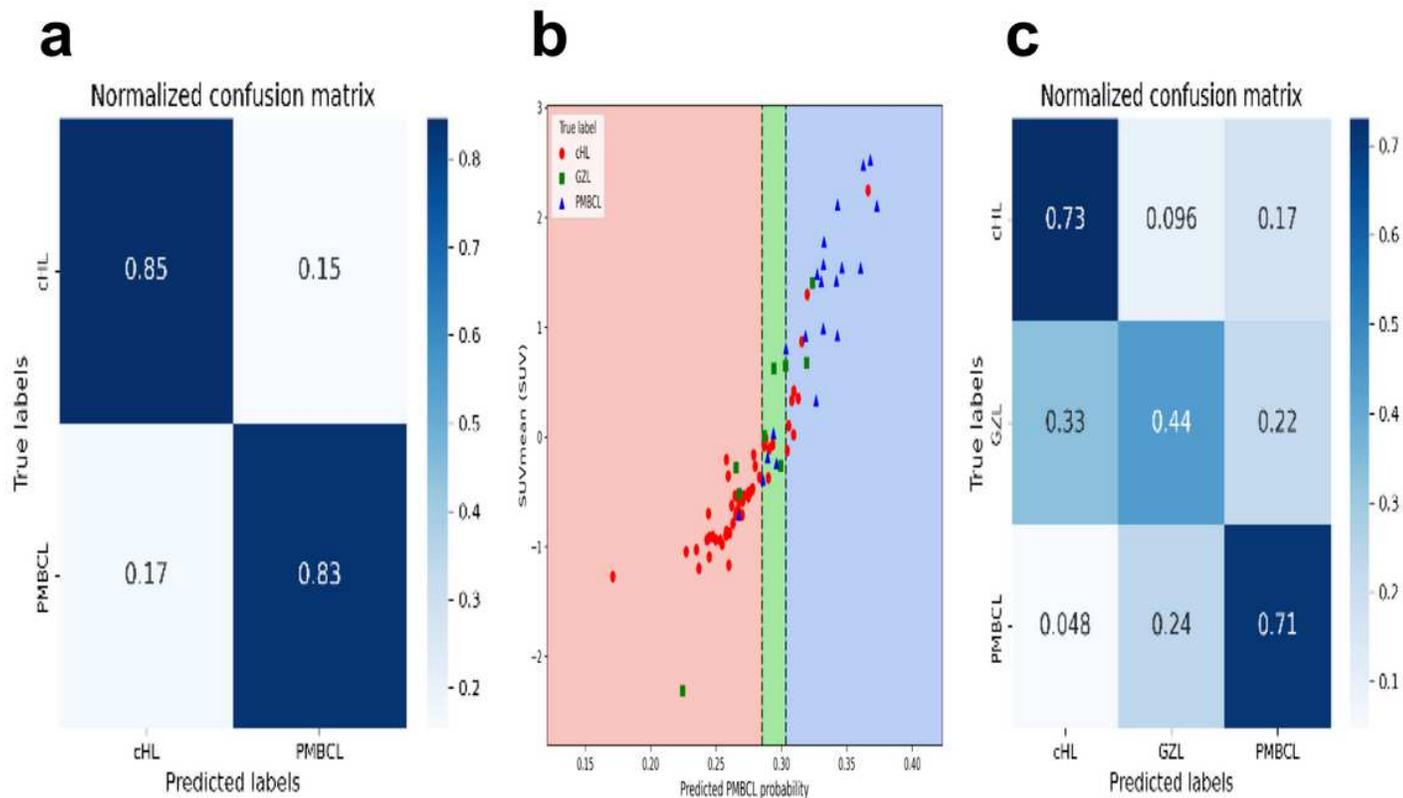


Figure 4

a) Normalized confusion matrix reporting the performance on the testing set (6 PMBCL and 13 cHL) of the Logistic Regression with custom boundary decision. The values reported on the diagonal are the TNR and TPR respectively (top-down), while the off-diagonal values are the FPR and FNR respectively (top-down). b) Training instances, with the addition of the GZL ones, are reported against the score achieved by the binary classifier. As expected on a histological basis, the GZL instances result in the middle of PMBCL-cHL classification. Opening an asymmetrical band around the binary decision boundary, it is possible to build three decision regions for multiclass predictions. c) Normalized confusion matrix reporting the performance on the training set (21 PMBCL, 52 cHL and 8 GZL) of the Logistic Regression with two custom decision boundaries. The values reported on the diagonal are the sensitivity with respect to the three labels: cHL, GZL and PMBCL respectively.