

Article

Automatic Prediction and Assessment of Treatment Response in Patients with Hodgkin's Lymphoma Using a Whole-Body DW-MRI Based Approach

Valentina Brancato ¹, Marco Aiello ^{1,*}, Roberta Della Pepa ², Luca Basso ¹, Nunzia Garbino ¹, Emanuele Nicolai ¹, Marco Picardi ², Marco Salvatore ¹ and Carlo Cavaliere ¹

- ¹ IRCCS SDN (Istituto di Ricovero e Cura a Carattere Scientifico, SYNLAB istituto di Diagnostica Nucleare), 80143 Napoli, Italy; valentina.brancato@synlab.it (V.B.); luca.basso@synlab.it (L.B.); nunzia.garbino@synlab.it (N.G.); emanuele.nicolai@synlab.it (E.N.); direzionescientifica@sdn-napoli.it (M.S.); carlo.cavaliere@synlab.it (C.C.)
- ² Department of Clinical Medicine and Surgery, Hematology Section, Federico II University of Naples, 80131 Naples, Italy; robertadellapepa@gmail.com (R.D.P.); marco.picardi@unina.it (M.P.)
- * Correspondence: marco.aiello@synlab.it

Received: 17 August 2020; Accepted: 15 September 2020; Published: 16 September 2020



Abstract: The lack of validation and standardization represents the main drawback for a clear role of whole-body diffusion weighted imaging (WB-DWI) for prediction and assessment of treatment response in Hodgkin's lymphoma (HL). We explored the reliability of an automatic approach based on the WB-DWI technique for prediction and assessment of response to treatment in patients with HL. The study included 20 HL patients, who had whole-body positron emission tomography (PET)/ magnetic resonance Imaging (MRI) performed before, during and after chemotherapy. Using the syngo.via MR Total Tumor Load tool, we automatically extracted values of diffusion volume (DV) and its associated histogram features by WB-DWI images, and evaluated their utility in predicting and assessing interim and end-of-treatment (EOT) response. The Mann–Whitney test followed by receiver operator characteristic (ROC) analysis was performed between features and their inter-time point percentage differences for patients having a complete or partial treatment response, revealing that several WB-DWI associated features allowed for prediction of interim response and both prediction and assessment of EOT response. Our proposed method offers huge advantages in terms of saving time and work, enabling clinicians to draw conclusions relating to HL treatment response in a fully automatic way, and encloses, also, all DWI advantages compared to PET/ computed tomography (CT).

Keywords: Hodgkin's Lymphoma; whole-body DWI; automatic tool; segmentation; response to treatment; ROC analysis

1. Introduction

Hodgkin's lymphoma (HL) is a relatively uncommon B-cell derived tumor, in which the unique cellular microenvironment is crucial for accurate diagnosis and pathobiology [1–3]. The role of diagnostic imaging provides important information for an accurate pretreatment evaluation and assessment of response to treatment, which are crucial steps for a good management of HL patients. In particular, the hybrid technique positron emission tomography (PET)/computed tomography (CT) with 18-fluorodeoxyglucose (¹⁸F-FDG) injection is considered the gold standard for HL management, from initial diagnosis to staging and assessment of response to treatment using the imaging-based Lugano classification [4–7].

Nevertheless, especially due to the harmful ionizing radiation dose involving both PET and CT modalities, there is an increasing interest towards integrated PET/MRI that combines the detailed



morphological information provided by the radiation-free MRI with the functional information that characterizes PET images. The excellent soft tissue contrast displayed by MRI, together with the high PET sensitivity, allows to perform robust diagnostic evaluations. The role of MRI is strengthened by the possibility to include functional MRI techniques, such as diffusion weighted imaging (DWI), in the protocol. DWI is a noninvasive tool that allows to quantify the random motion of water molecules (diffusion), which becomes hampered in structures characterized by high cellularity, such as lymphoma lesions. PET/CT, whole-body DWI (WB-DWI), as well, provides both anatomical and functional information. The characteristic DWI parameter is the apparent diffusion coefficient (ADC), which allows for a quantitative evaluation of changes in tissue cellularity, providing a useful tool for diagnosis and assessment of response to treatment in tumors, in particular for lymphomas. Several studies investigate the power of DWI for HL and Non-Hodgkin's (NHL) lymphoma diagnosis and assessment of treatment response compared to PET/CT, showing the potential role of DWI for these purposes [8].

Despite the continuous research and the promising results related to the usefulness of PET/MRI, especially when executed with DWI, the lack of validation and standardization represents the main drawback for a clear definition of the role of DWI in lymphoma diagnosis, staging, and response assessment. Moreover, it should be considered that, for a valid evaluation of response to treatment according to current guidelines, the readers need to have a deep experience and knowledge in the field of lymphomatous disease and their radiological evaluation [9]. Thus, we explore the possibility of having an automatic tool capable of providing support in lymphoma diagnosis.

MR Total Tumor Load stems from the promising results obtained by WB-DWI and ADC in multifocal disease, such as bone metastases and multiple myeloma, and the resulting need to dispose of an efficient tool able to evaluate this kind of lesions. This tool harnesses a threshold-based segmentation algorithm on whole-body diffusion-weighted images in order to identify regions of disease, and provides both the overall diffusion tumor volume and the histogram metrics of the corresponding computed ADC maps [10]. A high b-value image (acquired or computed) is used as input in order to maximize the contrast between lesions and healthy tissue. Several studies reported the advantages of using this tool especially in metastatic bone disease, but also in metastatic breast cancer and metastatic prostate cancer [10–13]. Another recent study (including three case studies) showed the benefits of the MR Total Tumor Load, not only in metastatic bone disease, but also in solid tumors. In particular, in one of these case studies concerning Hodgkin lymphoma in a 14-year-old girl, the reduction in tumor volume and the increase of low ADC values in ADC histogram between the pretreatment and the follow-up examination (after 2 months) are indicators of a good therapy response [14].

The purpose of this study is to evaluate the reliability of an automatic approach based on MR Total Tumor Load tool for WB-DWI technique relating to prediction and assessment of response to treatment in patients with HL, trying to detect the segmentation threshold, which is more capable of predicting and assessing response.

2. Materials and Methods

2.1. Patient Cohort

Twenty patients with histologically proven HL (11 men and 9 women), with a mean age of (35.7 ± 11.7) years were selected for this retrospective study. All subjects gave their informed consent for inclusion before they participated in the study. The study was conducted in accordance with the Declaration of Helsinki, and the protocol was approved by the Ethics Committee of the Istituto Nazionale Tumouri "Fondazione G. Pascale in 15 April 2020 (protocol number 3/20).

Inclusion criteria were patients being over the age of 18; histologic confirmation of HL at nodal biopsy; patients who underwent PET/CT followed by PET/MRI with WB-DWI at baseline before any treatment (T0), after two chemotherapy cycles (T1), and at the end-of-treatment (EOT)

(T2), acquired from February 2016 to July 2018. All patients received doxorubicin (also known as Adriamycin), bleomycin, vinblastine, dacarbazine (ABVD) chemotherapy, and were asked to complete medical history questionnaires and sign informed consents to undergo hybrid PET-CT and PET-MRI investigations. Characteristics of included patients are shown in Table 1.

Variable	Value
variable	value
No. of patients (<i>n</i>)	20
Age (y)	
Mean \pm SD	35.7 ± 11.7
Range	19-63
Gender $(n (\%))$	
Male	11 (55)
Female	9 (45)
HL Subtype (WHO classification) (n (%))	
Nodular sclerosis	13 (65)
Mixed cellularity	4 (20)
Lymphocyte rich	2 (10)
Lymphocyte depleted	1 (5)
B symptoms $(n \ (\%))$	
Fever	14 (70)
Sweats	9 (45)
Weight loss $\geq 10\%$	6 (30)
Histology $(n \ (\%))$	
Stage	
Ĩ	2 (10)
Ш	11 (55)
III	4 (20)
IV	3 (15)
Ervthrocyte sedimentation rate > 50 mm	2(10)

Table 1. Characteristics of study population.

Abbreviations: SD = Standard Deviation; HL = Hodgkin's Lymphoma; WHO = World Health Organization.

2.2. Acquisition Protocol

Data for all patients were acquired on both a PET-CT device and on a 3T hybrid PET-MRI system (Biograph mMR, Siemens, Erlangen, Germany) equipped with three 32-channel body coils, to cover the thorax, abdomen, and pelvis areas, and 12-channel phased array brain coils. Patients were asked to observe a fast of at least six hours. Sixty minutes after the 18-fluorodeoxyglucose (¹⁸F-FDG) injection by antecubital access, a PET-CT examination was performed from the brain vertex to the pelvis region. Then, patients underwent a whole-body PET-MRI protocol, which consisted of the following sequences: coronal T2 Turbo Inversion Recovery Magnitude (TIRM); an axial DWI sequence with b values of 50 and 800 s/mm²; axial and coronal T2 Half Fourier Acquisition Single Shot Turbo Spin Echo (HASTE), and an axial T1 Gradient Echo (GRE) in-out phase. The scan parameters are shown in Table 2. The attenuation correction is obtained from the segmentation (DIXON) into four classes, with predefined constant linear attenuation correction coefficients (LACs) for each class. The class denominations and the corresponding LACs were as follows: outer air (0 cm^{-1}) , lung (0.022 cm^{-1}) , fat tissue (0.085 cm⁻¹), soft tissue (0.1 cm⁻¹) [15]. Considering the PET data, the process of image reconstruction derived from an iterative algorithm called OSEM composed by three iterations on a matrix 172×172 . Moreover, dividing data into 21 subsets analyzed cyclically, it was possible to control the noise within low absorption regions.

Table 2. Scan parameters for whole-body MRI-DWI.

Sequence	Orientation	TR (ms)	TE (ms)	ST (mm)	TI (mm)	b-Values (s/mm ²)
T2 TIRM	Coronal	4500	84	5	220	
DWI	Axial		78	6	220	50,800
T2 HASTE	Axial and Coronal	1400	89	6		
T1 GRE	Axial	94	2.46	5		

Abbreviations: T2 TIRM = Turbo Inversion Recovery Magnitude; DWI = Diffusion-Weighted Imaging; T2 HASTE = Half Fourier Acquisition Single Shot Turbo Spin Echo; T1 GRE = Gradient Echo; TR: = Repetition Time; TE = Echo Time; ST = Slice Thickness; TI = Inversion Time.

2.3. PET Response Evaluation

A radiologist and a nuclear medicine physician, respectively, with 7 and 10 years of experience, assessed, by consensus response, to treatment examining PET images on the basis of the visual Deauville 5-point scale (5-PS), according to the Lugano classification criteria in two sessions. In the first session, the interim response to treatment was assessed evaluating PET images acquired at interim and comparing them to those acquired at baseline; in the second session, the EOT response was assessed evaluating PET images acquired at the EOT and comparing them to those acquired at baseline. Patients were classified as having a complete metabolic response (CMR) in case of 5-PS score of 1, 2, or 3 in lymph nodal and extra lymphatic sites with or without a residual mass and no evident FDG-uptake in marrow. Partial metabolic response (PMR) in case of 5-PS score of 4 or 5 with reduced uptake compared with baseline, residual lesions of any size, and, relating to bone marrow, residual uptake higher than in normal marrow, but reduced compared with baseline. Stable metabolic disease (SMD), in case of 5-PS score of 4 or 5 with no evident change in FDG uptake, or progressive metabolic disease (PMD), in case of 5-PS score of 4 or 5 in any lesion with an increase in intensity of FDG uptake from baseline, and/or new FDG-avid foci, consistent with lymphoma, as well as new or recurrent FDG-avid sites in bone marrow [4,7]. For cases with a Deauville assigned score of 4 or 5 at T2, PMR, SMD, or PMD was defined considering also the interim PET scan.

2.4. WB-DWI Image Analysis and Data Extraction

The analysis of WB-DWI images at each time point was performed using the syngo.via Frontier MR Total Tumor Load released research prototype v1.3.3 (Siemens Healthineers, Erlangen, Germany). For each patient, for each time point, the b800 images were used to automatically define threshold-based masks, using the threshold-based segmentation approach proposed by Blackedge et al. [16] and implemented in the syngo.via Frontier MR Total Tumor Load [10] (see Figure 1). Six segmentation threshold values were used, namely 5%, 10%, 20%, 40%, 60%, and 80%. No subsequent mask editing was made.



Figure 1. Whole-body diffusion weighted imaging (WB-DWI) image analysis and data extraction process using syngo.via Frontier MR Total Tumor Load software. Using the WB-DWI images as input (1. Data Loading step), b800 images were automatically segmented using setting a signal intensity threshold (e.g., 40%) at each WB-DWI acquisition time point (e.g., T0). No subsequent mask editing was made (2. Automatic segmentation). The overall mask volume (diffusion volume, DV) and the corresponding apparent diffusion coefficient (ADC) histogram metrics associated with the masked volume were extracted (3. Data extraction step). Yellow arrows link the three processing steps.

The overall mask volume (Diffusion Volume, DV) and the corresponding ADC histogram metrics associated with the masked volume were extracted. Specifically, extracted ADC histogram statistics were mean (ADCmean), standard deviation (ADCsd), median (ADCmd), 5% percentile (ADC5p), 95%

percentile (ADC95p), skewness (ADCsk), excess kurtosis (ADCkurt), entropy (ADCentr). The following notation was used in this study to indicate the single feature:

f^T_{th%},

where the superscript T indicates the time point at which WB-DWI images were acquired (T = 0, 1, 2), and the subscript th% indicates the segmentation threshold used (th = 5, 10, 20, 40, 60, 80). For each of the six thresholds, percentage changes in parameters during treatment from baseline and after treatment from baseline were calculated as follows:

$$\Delta f_{th\%}^{0T} = \frac{(f_{th\%}^{T} - f_{th\%}^{0})}{f_{th\%}^{0}} \times 100$$
(1)

where $f_{th\%}^{T}$ is the value of the feature at time point 1 or 2 and $f_{th\%}^{0}$ is the value of the feature at baseline. Percentage changes in parameters, after treatment from their values during treatment, were also calculated for each of the six thresholds as follows:

$$\Delta f_{\text{th}\%}^{12} = \frac{(f_{\text{th}\%}^2 - f_{\text{th}\%}^1)}{f_{\text{th}\%}^1} \times 100$$
⁽²⁾

where $f_{th\%}^1$ is the value of the feature at interim and $f_{th\%}^2$ is the value of the feature at the EOT.

2.5. Statistical Analysis

Values of each parameter were tested for normal distribution beforehand using a Kolmogorov–Smirnov test complemented by a graphical assessment for data normality using boxplots and Probability–Probability (P–P) plots, both overall and by subgroups identified, according to interim (T1) and EOT (T2) response. All normally distributed variables were summarized as mean (standard deviation), while those non-normally distributed as median and interquartile range (Q_1 ; Q_3).

In order to predict interim and EOT response, the *t*-test (in case of normally distributed parameters) or alternatively, the Mann–Whitney U test (in case of non-normally distributed parameters) on parameters at baseline, and also on parameters at interim for prediction of response at T2, was used to test the difference between CMR patients and each group of not CMR patients (PMR, SMD, PMD). For prediction of EOT response, percentage changes between features at T1 and T0 ($\Delta f_{th\%}^{01}$) was also evaluated. For parameters significant to t-test or Mann–Whitney test, receiver operator characteristic (ROC) curves were constructed and area under the curve (AUC) calculated to determine sensitivity and specificities and to find cut-off values that may be predictive of a poor response to treatment. The t-test or Mann–Whitney U test followed by ROC analysis for significant parameters was also performed on parameters at T1 and T2 as well as on percentage changes in parameters after treatment from baseline ($\Delta f_{th\%}^{02}$) and from interim ($\Delta f_{th\%}^{12}$) in order to evaluate their power in assessment of interim and EOT response.

Since none of the patients showed PMD or SMD at interim, and the population size for patients with PMD at the EOT was too small to perform statistical analysis (only two patients), we performed analysis for prediction of response to treatment only comparing patients with CMR and those with PMR. Specifically, we compared CMR and PMR patients at interim for prediction of interim response to treatment, and CMR and PMR patients at the EOT for prediction of EOT response.

All statistical analyses were performed using MATLAB (R2018a, MathWorks, Inc., Nettie, MA, USA). A *p*-value less than 0.05 was considered to indicate a statistically significant difference.

3. Results

3.1. Response to Therapy (Lugano Assessment)

Relating to the assessment of tumor response at interim, 14 patients (70%) showed CMR, while the remaining 6 patients (30%) had PMR. Relating to the assessment of tumor response at the EOT, 15 patients (75%) showed CMR, 3 (15%) had PMR, and 2 (10%) had PMD (see Table 3). In Figure 2, maximum intensity projection (MIP) of PET images at baseline, at interim and at the EOT for a patient showing CMR both at T1 and T2 are shown.

Table 3. Results of response to therapy according to Lugano assessment. Reported data are number of patients and percentages are in parenthesis.

Lugano Assessment	Response to Therapy at Interim (T1) (n (%))	Response to Therapy at the EOT (T2) (<i>n</i> (%))
CMR	14 (70)	15 (75)
PMR	6 (30)	3 (15)
SMD	0 (0)	0 (0)
PMD	0 (0)	2 (10)

Abbreviations: EOT = end of treatment; CMR = complete metabolic response; PMR = partial metabolic response; SMD = stable metabolic disease; PMD = progressive metabolic disease.



Figure 2. Maximum intensity projection (MIP) of positron emission tomography (PET) images at baseline, at interim and at the end-of-treatment (EOT) for a patient showing complete metabolic response (CMR), according to Lugano evaluation, both at T1 and T2.

3.2. Image Analysis

As shown in Figure 3, as the increase of the threshold, the tool became increasingly selective. Maximum intensity projection (MIP) of WB-DWI images acquired at b800 at baseline, at interim and at the EOT for a patient showing CMR both at T1 and T2 are also shown in Figure 3.





(c)

Figure 3. (a) Maximum intensity projection (MIP) of WB-DWI images at b800 at baseline (T0) and its relative masks at 5%, 10%, 20%, 40%, 60%, and 80%. (b) Maximum intensity projection (MIP) of whole-body DWI (WB-DWI) images at b800 at interim (T1) and its relative masks at 5%, 10%, 20%, 40%, 60%, and 80%. (c) Maximum intensity projection (MIP) of WB-DWI images at b800 at the EOT (T2) and its relative masks at 5%, 10%, 20%, 40%, 60%, and 80%. Images are from a patient showing complete metabolic response (according to Lugano evaluation) both at T1 and T2.

3.3. Prediction of Response to Treatment

The Kolmogorov–Smirnov test revealed that all baseline parameters, according to both T1 and T2 response, were non-Gaussian. Using Mann–Whitney U test, none of the baseline parameters was useful for prediction of interim response to therapy. Refer to Supplementary Table S1 for median values, interquartile ranges and associated *p*-values of all parameters, relating to prediction of interim response to treatment. Conversely, referring to prediction of EOT response, $DV_{40\%}^0$ and $DV_{60\%}^0$ were significantly higher in PMR patients than in CMR patients, while $ADC5p_{40\%}^0$ and $ADC95p_{20\%}^1$ were significantly lower in PMR patients than in CMR patients. Refer to Table 4 for median values, *p*-values and associated *p*-values of all parameters, relating for median values, interquartile ranges and associated *p*-values of all parameters. See Supplementary Materials for median values, interquartile ranges and associated *p*-values of all parameters, relating to prediction of EOT response (Tables S2 and S3) and for boxplots and ROC curves of significant features (Figures S1–S4).

8	of	13

Parameters ^a	Median (IQR)—CMR	Median (IQR)—PMR	p	AUC	Т	Sen (%)	Spec (%)	Acc (%)
$\frac{DV^{0}_{40\%}}{(cm^{3})}$	2033.6 (1907.5–2087.75)	2446.1 (2439.88–2542.55)	0.017	0.93	2342.15	100	93	94
$DV_{60\%}^{0}$ (cm ³)	1525.7 (1440.48–1556.23)	1667.4 (1626.23–1733.4)	0.01	0.96	1604.6	100	93	94
ADC5p ⁰ _{40%} (×10 ⁻⁶ mm ² /s)	634.5 (607.5–634.5)	580.5 (540–583.88)	0.037	0.89	596.25	80	100	83
ADC95p ¹ _{20%} (×10 ⁻⁶ mm ² /s)	2686.5 (2416.5–2902.5)	2335.5 (2254.5–2376)	0.032	0.91	2403	85	100	88

Table 4. Significant results for prediction of end-of-treatment response.

^a superscripts indicate WB-DWI acquisition time point; subscripts indicate segmentation threshold. IQR = Interquartile Range; CMR = Complete Metabolic Response; PMR = Partial Metabolic Response; p = p-value; AUC = Area under the ROC Curve; T = Optimal Threshold; Sen = Sensitivity; Spec = Specificity; Acc = Accuracy; DV = Diffusion Volume; ADC = Apparent Diffusion Coefficient; ADC5p = ADC 5% percentile; ADC95p = ADC 95% percentile. Units of measurement for each parameter is shown in square brackets.

3.4. Assessment of Response to Treatment

Kolmogorov-Smirnov test revealed that all interim and EOT parameters, according to both T1 and T2 response, were non-Gaussian. Using Mann-Whitney U test for assessment of interim response to treatment, values for ADCmean¹_{20%}, ADCsd¹_{20%}, and ADC95p¹ at 20% and 40% were significantly lower in PMR patients than in CMR patients. For thresholds from 5% to 40%, Δ ADCsd⁰¹ was found to be significantly higher in CMR patients than in PMR patients. The same trend was observed for Δ ADC95p⁰¹ at threshold of 5%, 20%, 40% and 60% and for Δ ADCentr⁰¹ at thresholds from 5% to 40%. Conversely, values at interim of DV at 40% showed a completely opposite trend. In Table 5, median values and associated ROC analysis statistics for discrimination between CMR and PMR patients are reported. See Supplementary Tables S4 and S5 for median values, interquartile ranges, and associated *p*-values of all parameters relating to assessment of response to treatment at interim, and Figures S5–S22 for boxplots and ROC curves of significant features.

Table 5. Significant results for assessment of interim response to treatment.

Parameters ^a	Median (IQR)—CMR	Median (IQR)—PMR	р	AUC	Т	Sen (%)	Spec (%)	Acc (%)
${ m DV}^1_{40\%}$ (cm ³)	1916.35 (1846.85–2084.55)	2311.45 (2235.5–2446.7)	0.005	0.90	2116.5	100	83	89
ADCmean ¹ _{20%} (×10 ⁻⁶ mm ² /s)	1311.58 (1252.74–1356.32)	1221.34 (1111.06–1240.01)	0.009	0.88	1243.85	83	83	83
ADCsd ¹ _{20%} (×10 ⁻⁶ mm ² /s)	695.49 (644.57–737.93)	609.96 (577.22–635.21)	0.009	0.88	656.93	75	100	83
ADC95p ¹ _{20%} (×10 ⁻⁶ mm ² /s)	2686.5 (2497.5–2902.5)	2281.5 (2227.5–2389.5)	0.001	0.93	2403	92	100	94
ADC95p ¹ _{40%} (×10 ⁻⁶ mm ² /s)	2646 (2349–2796.75)	2214 (2093–2335.5)	0.013	0.86	2362.5	75	83	78
ADCentr $^{1}_{20\%}$ (×10 ⁻⁶ mm ² /s)	8.49 (8.4–8.54)	8.37 (8.3–8.47)	0.024	0.83	8.48	67	100	78
$\begin{array}{c} \Delta \text{ADCsd}_{5\%}^{01} \\ (\%) \end{array}$	4.85 (1.87–5.66)	-4.79 (-9.43 to -2.67)	0.006	0.92	0.81	83	100	88
$\begin{array}{c}\Delta \text{ADCsd}_{10\%}^{01}\\(\%)\end{array}$	2.69 (-0.08 to 5.12)	-7.38 (-10.08 to -0.04)	0.036	0.83	0.1	75	80	76
$\begin{array}{c} \Delta \text{ADCsd}_{20\%}^{01} \\ (\%) \end{array}$	8.47 (-2.99 to 12.54)	-10.37 (-14.92 to -5.93)	0.013	0.88	-6.79	83	80	82
$\frac{\Delta \text{ADCsd}_{40\%}^{01}}{(\%)}$	7.76 (-0.09 to 12)	-9.52 (-21.81 to -2.84)	0.019	0.87	-2.05	83	100	88
(%)	(-1.41 to 3.49)	-2.72 (-8.86 to -2.36)	0.045	0.82	-2.09	0.83	0.8	0.82
$\Delta ADC95p_{20\%}^{01}$ (%)	2.8 (-5.56 to 6.43)	-18.43 (-20.71 to -6.46)	0.013	0.88	-5.52	75	80	76
$\Delta ADC95p_{40\%}^{01}$ (%)	4.65 (-3.64 to 13.14)	-10.33 (-24.45 to -1.82)	0.02	0.86	-1.21	75	80	76
$\Delta ADC95p_{60\%}^{01}$ (%)	1.95 (-4.34 to 6.81)	-3.85 (-15.68 to 0.31)	0.048	0.82	1.28	0.75	1	0.82

Parameters ^a	Median (IQR)—CMR	Median (IQR)—PMR	р	AUC	Т	Sen (%)	Spec (%)	Acc (%)
$\Delta ADCentr_{5\%}^{01}$ (%)	0.32 (-0.04 to 0.56)	-0.53 (-1.24 to -0.19)	0.013	0.88	-0.27	92	8	88
$\begin{array}{c} \Delta \text{ADCentr}_{10\%}^{01} \\ (\%) \end{array}$	0.27 (0.16–0.56)	-0.51 (-1.16 to -0.03)	0.019	0.87	0.02	83	80	82
$\Delta ADCentr_{20\%}^{01}$ (%)	0.71 (-0.27 to 1.21)	-0.83 (-2.16 to -0.39)	0.009	0.90	-0.03	75	100	82
$\begin{array}{c} \Delta \text{ADCentr}^{01}_{40\%} \\ (\%) \end{array}$	0.88 (-0.18 to 1.36)	-1.42 (-1.65 to -0.92)	0.013	0.88	-1.15	92	80	88

Table 5. Cont.

^a superscripts indicate one or two (in case of percentage differences) WB-DWI acquisition time points; subscripts indicate segmentation threshold. IQR = Interquartile Range; CMR = Complete Metabolic Response; PMR = Partial Metabolic Response; p = p-value; AUC = Area Under the ROC Curve; T = Optimal Threshold; Sen = Sensitivity; Spec = Specificity; DV = Diffusion Volume; ADC = Apparent Diffusion Coefficient; ADCmean = ADC mean; ADCsd = ADC standard deviation; ADC95p = ADC 95% percentile; ADCentr = ADC entropy; Δ ADC = percentage change of ADC between two time points; Δ ADCsd = percentage change in ADC 95% percentile; Δ ADCentr

On the other hand, relating to assessment of EOT response, values at T2 of ADC mean and median at 5% threshold, $\Delta ADCsd_{5\%}^{02}$, $\Delta ADC5p_{60\%}^{02}$, and $\Delta ADCsd_{5\%}^{12}$ were significantly higher in PMR than in CMR patients. Conversely, $\Delta DV_{40\%}^{02}$ and $\Delta ADCsd_{60\%}^{02}$ revealed a completely opposite trend. Refer to Table 6 for median values, *p*-values, and associated ROC analysis statistics for discrimination between CMR and PMR patients. See Supplementary Materials for median values, interquartile ranges, and associated *p*-values of all parameters, relating to assessment of response to treatment at interim (Tables S6 and S7) and for boxplots and ROC curves of significant features (Figures S23–S29).

Table 6. Significant results for assessment of end-of-treatment response.

Parameters ^a	Median (IQR)—CMR	Median (IQR)—PMR	p	AUC	Т	Sen (%)	Spec (%)	Acc (%)
$\begin{array}{l} \text{ADCmean}_{5\%}^2 \\ \text{($\times10^{-6} \text{ mm}^2/\text{s}$)} \end{array}$	1103.54 (1061.37–1159.35)	1211.08 (1180.53–1227.21)	0.014	0.95	1169.09	100	85	88
$ADCmd_{5\%}^2$ (×10 ⁻⁶ mm ² /s)	1066.5 (1034.44–1174.5)	1228.5 (1208.25–1289.25)	0.01	0.96	1188	100	92	94
$\Delta DV^{02}_{40\%}$ (%)	15.04 (4.49–34.57)	-10.05 (-22.72 to -1.69)	0.025	0.92	4.06	77	100	81
$\Delta ext{ADCsd}_{5\%}^{02}$ (%)	-2.09 (-4.8 to 2.38)	4.6 (2.99–4.91)	0.039	0.9	2.36	100	77	81
$\begin{array}{c} \Delta \mathrm{ADCsd}^{02}_{60\%} \\ (\%) \end{array}$	-0.84 (-5.82 to 4.66)	-23.12 (-28.21 to -10.25)	0.025	0.92	-5.79	77	100	81
$\Delta ADC5p_{60\%}^{02}$ (%)	-4.26 (-6.69 to 0.91)	4.26 (4.26–7.26)	0.014	0.95	4.17	100	92	94
$\begin{array}{c} \Delta \mathrm{ADCsd}_{5\%}^{12} \\ (\%) \end{array}$	-1.79 (-4.5 to 1.22)	6.35 (5.09–10.09)	0.017	0.94	4.21	100	92	93

^a superscripts indicate one or two (in case of percentage differences) DWI acquisition time points; subscripts indicate segmentation threshold. IQR = Interquartile Range; CMR = Complete Metabolic Response; PMR = Partial Metabolic Response; p = p-value; AUC = Area Under the ROC Curve; T = Optimal Threshold; Sen = Sensitivity; Spec = Specificity; DV = Diffusion Volume; ADC = Apparent Diffusion Coefficient; ADCmean = ADC mean; ADCmd = ADC median; Δ ADCsd = percentage change in ADC standard deviation; Δ ADC5p = percentage change in ADC 5% percentile.

4. Discussion

The lack of validation and standardization represents the main drawback for a clear definition of the role of WB-DWI in lymphoma diagnosis, staging, and response assessment [17–20]. In the present preliminary study, we investigated a new WB-DWI-based approach for prediction and assessment of lymphoma response to treatment through the analysis of quantitative WB-DWI histogram features extracted from MR Total Tumor Load tool at six segmentation thresholds (5%, 10%, 20%, 40%, 60%, 80%) and using Lugano criteria applied on PET/CT images as reference standard.

10 of 13

Results of our study revealed the inability of all examined parameters to predict interim response. Concerning prediction of EOT response, DV at 40% and 60% was found to be significantly higher in PMR patients than in CMR patients. This could be related to the poorer response to treatment of PMR patients, that is normally associated with a higher diffusion tumor volume associated to DWI signal intensity [14,17]. Relating to ADC histogram variables, values at baseline of the 5th percentile of ADC at 40% and those of ADC skewness at 20% were respectively significantly lower and higher in PMR than in CMR patients and were found to be able to predict EOT response. Interim value of 95th percentile of ADC at 20% was also found to be useful for this purpose.

Interim value of DV at 40% was also useful for the assessment of interim response, as well as the following histogram related features: values at interim of ADC mean, standard deviation, and entropy at 20% threshold, and values of ADC 95th percentile at 5%, 20%, 40%, and 60%, were significantly lower in PMR than in CMR. The above-mentioned findings on ADC mean are consistent with the definition of PMR patients (which should have an overall lower ADC mean value than CMR patients) and in line with those obtained in the case study reported by Tsiflikas et al. [14] on a 14-year girl with HL, as well as in previous studies involving other tumor types and using MR Total Tumor Load Tool [4,10,12]. Percentage change between T1 and T0 in ADC standard deviation at 5%, 10%, 20%, and 40% was significantly lower in PMR patients than in CMR patients, and it was due to a completely opposite trend: in PMR patients this feature increased from T0 to T1. Same behavior was observed for percentage change between T1 and T0 in ADC 95th percentile at 20% and 40%, and in ADC entropy from 5% to 40% threshold.

Concerning the assessment of EOT response, mean and median values of ADC at the EOT were found to be significantly higher in PMR than in CMR patients. Significant results for assessment of EOT response were also achieved by percentage change between T2 and T0 in DV at 40% and in ADC 5th percentile at 60%, and by percentage change between T2 and T1 in ADC standard deviation at 5%.

Obtained results support the theory that DV and its related histogram-based ADC statistics could be useful in prediction and assessments of HL response to therapy, as observed in previous mentioned studies concerning different oncologic diseases and using MR Total Tumor Load Tool [10,12,14].

However, our results cannot be directly compared with any of these, since in our study, DV and its relative ADC histogram parameters were associated to a diffusion volume mask automatically generated by the tool, and so not manually segmented.

In fact, the innovation in our research is that we tried to draw conclusions skipping the mask editing step expected by the tool, which would require the intervention of an expert operator able to exclude normally hyperintense and not tumoral regions, and directly extract the features associated to the unrefined mask.

Our proposed method would surely offer huge advantages in terms of saving time and work, enabling also less expert operator to draw conclusions relating to lymphoma diagnosis. Moreover, being based on WB-DWI technique, it also encloses all DWI advantages compared to CT and PET, such as the absence of ionizing radiation, the fast acquisition of images and the no-requirement for contrast injection [21,22]. However, at the same time, it adds substantial limitations to our study. First of all, artifact regions and/or normal hyperintense regions were incorporated in the automatically segmented mask. For example, organs such as brain, kidneys, and spleen, are usually hyperintense. Moreover, due to its frequent activation due to chemotherapy, the bone marrow also appeared hyperintense, and this could justify an increase in diffusion volume during the chemotherapeutic treatment.

Limitations related to the gold standard are that the 5-PS and the subsequent Lugano assessment are a qualitative gold standard based on visual assessment, which results influenced by inter-observer variability due to the subjectivity of the interpretation [23]. We could have been accompanied this qualitative assessment with a semiquantitative one based on Standardized Uptake Value (SUV) values.

In order to mitigate this limitation, Lugano assessment was performed in consensus between radiologist and nuclear medicine physician.

Furthermore, variations in time interval between the three image acquisitions and the different stage at T0 among patients may have influenced results.

Finally, the patient sample involved was small and unbalanced, and this might have adversely affected statistical results. Moreover, it should be considered that the retrospective study nature characterizing our study is supposed to have more bias and should be validated through prospective studies [24]. Despite this, the novelty of our method may provide a basis for future retrospective and prospective studies involving more participants. More informative results may be obtained if more patients could be investigated, also considering that, only three patients were found to have PMR response at EOT, making results on prediction and assessment of EOT response imprecise and unreliable. Moreover, only two patients were found to have PMD and no one SMD, not allowing us to perform comparisons among these categories.

It could be interesting to evaluate PET images using the same method, and compare them with WB-DWI images, both in terms of histogram parameters and relating to masked total volume, respectively, total diffusion volume (DV) in WB-DWI and total metabolic volume in PET), and investigate how unrefined diffusion volume masks extracted by the tool are related with SUV values associated with PET of PET/MRI. Following this line, it would be attractive to investigate the utility of the unrefined total metabolic volume automatically extracted by the tool for prediction and assessment of response to treatment in lymphoma patients, as done by Cottereau et al. [25,26] considering the total metabolic tumor volume (TMTV). Moreover, it would be interesting to integrate imaging information with those arising from lymph node biopsy (which is the gold standard for diagnosing lymphoma) and also from liquid biopsy markers, in order to choose a tailored treatment strategy for each HL patient and better evaluate treatment efficacy [27–29].

5. Conclusions

In conclusion, in this preliminary study, we found that several WB-DWI associated features allowed for prediction of interim response and both prediction and assessment of EOT response of patients with HL. However, the novelty of our method of feature extraction, with its related restrictions, and the lack of a defined and standardized role of DWI for the management of HL, pave the way for further studies involving larger groups of patients, which are essential to investigate the effective impact of our method and validate obtained results.

Supplementary Materials: The following are available online at http://www.mdpi.com/2075-4418/10/9/702/s1, Tables S1–S7, Figures S1–S29.

Author Contributions: C.C. and V.B. conceived the study; R.D.P., M.P., E.N., L.B., and N.G. enrolled the patients and acquired data; V.B., L.B., and N.G. performed data analysis; C.C., M.A., M.P., and V.B. contributed to the interpretation of the results; V.B. wrote the original draft; C.C., M.S., and M.P. reviewed and edited the manuscript; M.S. and C.C. supervised the project. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by "Ricerca Corrente" Grant from Italian Ministry of Health (IRCCS SDN).

Conflicts of Interest: The authors declare no conflict of interest.

References

- Vadakara, J.; Andrick, B. Current advances in Hodgkin's lymphoma. *Chronic Dis. Transl. Med.* 2019, 5, 15–24. [CrossRef] [PubMed]
- 2. Wang, H.-W.; Balakrishna, J.P.; Pittaluga, S.; Jaffe, E.S. Diagnosis of Hodgkin lymphoma in the modern era. *Br. J. Haematol.* **2018**, *184*, 45–59. [CrossRef] [PubMed]
- 3. Bröckelmann, P.J.; Eichenauer, D.A.; Jakob, T.; Follmann, M.; Engert, A.; Skoetz, N. Hodgkin Lymphoma in Adults. *Dtsch. Aerzteblatt Online* **2018**, *115*, 535–540. [CrossRef]

- 4. Cheson, B.D.; Fisher, R.I.; Barrington, S.; Cavalli, F.; Schwartz, L.H.; Zucca, E.; Lister, T.A. Recommendations for Initial Evaluation, Staging, and Response Assessment of Hodgkin and Non-Hodgkin Lymphoma: The Lugano Classification. *J. Clin. Oncol.* **2014**, *32*, 3059–3067. [CrossRef] [PubMed]
- 5. Wang, X. PET/CT: Appropriate application in lymphoma. Chin. Clin. Oncol. 2015, 4, 4.
- 6. Kulkarni, N.M.; Pinho, D.F.; Narayanan, S.; Kambadakone, A.R.; Abramson, J.S.; Sahani, D.V. Imaging for Oncologic Response Assessment in Lymphoma. *Am. J. Roentgenol.* **2017**, *208*, 18–31. [CrossRef]
- Cheson, B.D. Staging and response assessment in lymphomas: The new Lugano classification. *Chin. Clin. Oncol.* 2015, 4, 9.
- 8. Fitzpatrick, J.J.; Ryan, M.A.; Bruzzi, J.F. Diagnostic accuracy of diffusion-weighted imaging- magnetic resonance imaging compared to positron emission tomography/computed tomography in evaluating and assessing pathological response to treatment in adult patients with lymphoma: A systematic review. *J. Med. Imaging Radiat. Oncol.* **2018**, *62*, 530–539. [CrossRef]
- 9. Van Heertum, R.L.; Scarimbolo, R.; Wolodzko, J.G.; Klencke, B.; Messmann, R.; Tunc, F.; Sokol, L.; Agarwal, R.; Strafaci, J.A.; O'Neal, M. Lugano 2014 criteria for assessing FDG-PET/CT in lymphoma: An operational approach for clinical trials. *Drug Des. Dev. Ther.* **2017**, *11*, 1719–1728. [CrossRef]
- 10. Grimm, R.; Padhani, A.R. Whole-body Diffusion-weighted MR Image Analysis with syngo.via Frontier MR Total Tumor. *Magn. Flash* **2017**, *68*, 73–75.
- 11. Padhani, A.R. Observing Endocrine Therapy Resistance in Metastatic Breast Cancer with Whole-body MRI. *Magn. Flash* **2017**, *68*, 80–83.
- 12. Dalili, D.; Padhani, A.R.; Grimm, R. Quantitative WB-MRI with ADC Histogram Analysis for Response Assessment in Diffuse Bone Disease. *Magn. Flash* **2017**, *69*, 32–37.
- 13. Padhani, A.R.; Tunariu, N. Metastatic Prostate Cancer in Practice—The MET-RADS-P Imaging Response System Using Whole-body MRI. *Magn. Flash* **2017**, *68*, 64–72.
- 14. Tsiflikas, I. MR Total Tumor Load—First Clinical Experience in Pediatric Oncology Patients. *Magn. Flash* **2019**, *73*, 72–76.
- 15. Martinez-Möller, A.; Souvatzoglou, M.; Delso, G.; Bundschuh, R.A.; Chefd'Hotel, C.; Ziegler, S.; Navab, N.; Schwaiger, M.; Nekolla, S.G. Tissue Classification as a Potential Approach for Attenuation Correction in Whole-Body PET/MRI: Evaluation with PET/CT Data. *J. Nucl. Med.* **2009**, *50*, 520–526. [CrossRef] [PubMed]
- Blackledge, M.; Collins, D.J.; Tunariu, N.; Orton, M.R.; Padhani, A.R.; Leach, M.O.; Koh, D.-M. Assessment of Treatment Response by Total Tumor Volume and Global Apparent Diffusion Coefficient Using Diffusion-Weighted MRI in Patients with Metastatic Bone Disease: A Feasibility Study. *PLoS ONE* 2014, 9, e91779. [CrossRef]
- Toledano-Massiah, S.; Luciani, A.; Itti, E.; Zerbib, P.; Vignaud, A.; Belhadj, K.; Baranes, L.; Haioun, C.; Lin, C.; Rahmouni, A. Whole-Body Diffusion-weighted Imaging in Hodgkin Lymphoma and Diffuse Large B-Cell Lymphoma. *Radiographics* 2015, *35*, 747–764. [CrossRef] [PubMed]
- Littooij, A.S.; Kwee, T.C.; De Keizer, B.; Bruin, M.C.; Coma, A.; Beek, F.J.; Fijnheer, R.; Nievelstein, R.A.J. Whole-body MRI-DWI for assessment of residual disease after completion of therapy in lymphoma: A prospective multicenter study. *J. Magn. Reson. Imaging* 2015, 42, 1646–1655. [CrossRef]
- 19. Stecco, A.; Buemi, F.; Iannessi, A.; Carriero, A.; Gallamini, A. Current concepts in tumor imaging with whole-body MRI with diffusion imaging (WB-MRI-DWI) in multiple myeloma and lymphoma. *Leuk. Lymphoma* **2018**, *59*, 2546–2556. [CrossRef] [PubMed]
- 20. Kwee, T.C.; Basu, S.; Torigian, E.A.; Nievelstein, R.A.J.; Alavi, A. Evolving Importance of Diffusion-Weighted Magnetic Resonance Imaging in Lymphoma. *PET Clin.* **2012**, *7*, 73–82. [CrossRef]
- 21. Koh, D.-M.; Collins, D.J. Diffusion-Weighted MRI in the Body: Applications and Challenges in Oncology. *Am. J. Roentgenol.* **2007**, *188*, 1622–1635. [CrossRef] [PubMed]
- 22. Kwee, T.C.; Takahara, T.; Ochiai, R.; Koh, D.-M.; Ohno, Y.; Nakanishi, K.; Niwa, T.; Chenevert, T.L.; Luijten, P.R.; Alavi, A. Complementary Roles of Whole-Body Diffusion-Weighted MRI and 18F-FDG PET: The State of the Art and Potential Applications. *J. Nucl. Med.* **2010**, *51*, 1549–1558. [CrossRef] [PubMed]
- 23. Sadik, M.; Lind, E.; Polymeri, E.; Enqvist, O.; Ulén, J.; Trägårdh, E. Automated quantification of reference levels in liver and mediastinal blood pool for the Deauville therapy response classification using FDG-PET/CT in Hodgkin and non-Hodgkin lymphomas. *Clin. Physiol. Funct. Imaging* **2018**, *39*, 78–84. [CrossRef] [PubMed]
- 24. Norvell, D.C. Study types and bias—Don't judge a study by the abstract's conclusion alone. *Evid. Based Spine-Care J.* **2010**, *1*, 7–10. [CrossRef] [PubMed]

- 25. Cottereau, A.-S.; El-Galaly, T.C.; Becker, S.; Broussais, F.; Petersen, L.J.; Bonnet, C.; Prior, J.O.; Tilly, H.; Hutchings, M.: Casaspause, Q.: et al. Predictive Value of PET Response Combined with Baseline Metabolic
- Hutchings, M.; Casasnovas, O.; et al. Predictive Value of PET Response Combined with Baseline Metabolic Tumor Volume in Peripheral T-Cell Lymphoma Patients. *J. Nucl. Med.* **2017**, *59*, 589–595. [CrossRef]
- Cottereau, A.-S.; Hapdey, S.; Chartier, L.; Modzelewski, R.; Casasnovas, O.; Itti, E.; Tilly, H.; Vera, P.; Meignan, M.; Becker, S. Baseline Total Metabolic Tumor Volume Measured with Fixed or Different Adaptive Thresholding Methods Equally Predicts Outcome in Peripheral T Cell Lymphoma. J. Nucl. Med. 2016, 58, 276–281. [CrossRef]
- 27. Moia, R.; Favini, C.; Rasi, S.; Deambrogi, C.; Ferri, V.; Schipani, M.; Sagiraju, S.; Mahmoud, A.M.; Kodipad, A.A.; Adhinaveni, R.; et al. Liquid biopsy in lymphomas: A potential tool for refining diagnosis and disease monitoring. *J. Cancer Metastasis Treat.* **2019**, *2019*. [CrossRef]
- 28. Rossi, D.; Spina, V.; Bruscaggin, A.; Gaidano, G. Liquid biopsy in lymphoma. *Haematologica* **2019**, *104*, 648–652. [CrossRef]
- 29. Zanfardino, M.; Franzese, M.; Pane, K.; Cavaliere, C.; Monti, S.; Esposito, G.; Salvatore, M.; Aiello, M. Bringing radiomics into a multi-omics framework for a comprehensive genotype-phenotype characterization of oncological diseases. *J. Transl. Med.* **2019**, *17*, 337. [CrossRef]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).