RESEARCH ARTICLE

Utility of diffusion MRI characteristics of cervical lymph nodes as disease classifier between patients with head and neck squamous cell carcinoma and healthy volunteers

Marianthi-Vasiliki Papoutsaki[1] | Harbir Singh Sidhu[1] | Nikolaos Dikaios[2] |
Saurabh Singh[1] | David Atkinson[1] | Baris Kanber[3] | Timothy Beale[4] |
Simon Morley[4] | Martin Forster[5,6] | Dawn Carnell[6] | Ruheena Mendes[6] |
Shonit Punwani[1]

[1]Centre for Medical Imaging, University College London, London, UK

[2]Centre for Vision, Speech and Signal Processing, University of Surrey, Guildford, UK

[3]Centre for Medical Image Computing, Department of Medical Physics and Biomedical Engineering, University College London, London, UK

[4]Department of Radiology, University College London Hospital, London, UK

[5]Department of Oncology, University College London, Cancer Institute, London, UK

[6]Department of Oncology, University College London Hospital, London, UK

Correspondence
Shonit Punwani, Centre for Medical Imaging, University College London, 2nd Floor, Charles Bell House, 43-45 Foley Street, London W1W 7TS, UK.
Email: shonit.punwani@gmail.com

Diffusion MRI characteristics assessed by apparent diffusion coefficient (ADC) histogram analysis in head and neck squamous cell carcinoma (HNSCC) have been reported as helpful in classifying tumours based on diffusion characteristics. There is little reported on HNSCC lymph nodes classification by diffusion characteristics. The aim of this study was to determine whether pretreatment nodal microstructural diffusion MRI characteristics can classify diseased nodes of patients with HNSCC from normal nodes of healthy volunteers. Seventy-nine patients with histologically confirmed HNSCC prior to chemoradiotherapy, and eight healthy volunteers, underwent diffusion-weighted (DW) MRI at a 1.5-T MR scanner. Two radiologists contoured lymph nodes on DW (b = 300 s/m^2) images. ADC, distributed diffusion coefficient (DDC) and alpha (α) values were calculated by monoexponential and stretched exponential models. Histogram analysis metrics of drawn volume were compared between patients and volunteers using a Mann–Whitney test. The classification performance of each metric between the normal and diseased nodes was determined by receiver operating characteristic (ROC) analysis. Intraclass correlation coefficients determined interobserver reproducibility of each metric based on differently drawn ROIs by two radiologists. Sixty cancerous and 40 normal nodes were analysed. ADC histogram

Abbreviations used: ADC, apparent diffusion coefficient; AUC, area under the curve; BoT, base of tongue; CRT, chemoradiotherapy; CT, computed tomography; DDC, distributed diffusion coefficient; DW, diffusion-weighted; FNA, fine needle aspiration; GTS, glossotonsillar sulcus; HNSCC, head and neck squamous cell carcinoma; ICC, intraclass correlation coefficient; NPV, negative predictive value; PPV, positive predictive value; PROPELLER, Periodically Rotated Overlapping ParallEL Lines with Enhanced Reconstruction; ROC, receiver operating characteristic; ROI, region of interest; UICC, Union for International Cancer Control; US, ultrasound.

analysis revealed significant differences between patients and volunteers ($p \leq 0.0001$ to 0.0046), presenting ADC distributions that were more skewed (1.49 for patients, 1.03 for volunteers; $p = 0.0114$) and 'peaked' (6.82 for patients, 4.20 for volunteers; $p = 0.0021$) in patients. Maximum ADC values exhibited the highest area under the curve ([AUC] 0.892). Significant differences were revealed between patients and volunteers for DDC and α value histogram metrics ($p \leq 0.0001$ to 0.0044); the highest AUC were exhibited by maximum DDC (0.772) and the 25th percentile α value (0.761). Interobserver repeatability was excellent for mean ADC (ICC = 0.88) and the 25th percentile α value (ICC = 0.78), but poor for all other metrics. These results suggest that pretreatment microstructural diffusion MRI characteristics in lymph nodes, assessed by ADC and α value histogram analysis, can identify nodal disease.

KEYWORDS
apparent diffusion coefficient, diffusion magnetic resonance imaging, head and neck squamous cell carcinoma, lymph nodes

1 | INTRODUCTION

Computed tomography (CT) and magnetic resonance imaging (MRI) have similar performances for the detection of cervical nodal metastasis, with a recently reported accuracy of 90% and 94%, respectively.[1] Local disease recurrence following therapy can occur in 30% to 40% of patients with head and neck squamous cell carcinoma (HNSCC) who present with nodal metastases at initial staging.[1] One possible explanation for local recurrence may be the inaccuracy of detecting all disease sites using conventional imaging, resulting in undertreatment of undetected disease sites.

Diffusion-weighted (DW) MRI and the assessment of apparent diffusion coefficient (ADC) are considered powerful imaging tools in oncology.[2] DW-MRI assesses the diffusion of water molecules within biological tissues and correlates it to the tissue cellularity and the integrity of cell membranes, thereby reflecting the microstructure of tissue. ADC is conventionally derived by fitting a monoexponential model to the signal (S) acquired across the images with different degrees of sensitivity to water motion, that is, different b-value[3]:

$$S(b) = S_0 e^{-(b \times ADC)} \tag{1}$$

The last decade has seen increased exploration of ADC as an imaging biomarker for the detection[4,5] of HNSCC. However, research reports varying degrees of success when using DW-MRI (with simple calculation of monoexponential ADC) for this purpose.[6] This is due to the fact that ADC is a summative parameter that gives a single value for all microstructures within a given voxel. It does not provide a breakdown across different sources of diffusion signal weighting (e.g. vascular flow, extracellular water, intracellular water) and hence is a broad-brush representation of all diffusion processes occurring within a voxel. To overcome this, a range of nonmonoexponential models has been developed that can extract additional parameters from DW images, providing more specific information of individual diffusion components. For example, at the simplest level, bi-exponential diffusion models allow the separation of the perfusion and diffusion effects contributing to the diffusion signal.[7] Another approach (which acknowledges that specific separation of diffusion signals into different microstructural elements is unlikely to be precise) has been to assess the overall heterogeneity of diffusion rates within individual voxels through a stretched exponential model. This model has been previously applied preclinically[8] and clinically.[9] It estimates the distributed diffusion coefficient (DDC) and an additional parameter, the alpha (α) value related to the intravoxel heterogeneity of diffusion coefficients:

$$S(b) = S_0 \left(e^{(-b \times DDC)^{\alpha}} \right) \tag{2}$$

Heterogeneity has been suggested as an important parameter in predicting the outcomes of patients with cancer.[10] Structural heterogeneity on CT has been assessed using textural analysis techniques and has also demonstrated a similar relationship with outcomes.[11] Diffusion heterogeneity assessed by histogram analysis in HNSCC patients has previously been reported as helpful in classifying primary tumour sites.[12,13] To date, little has been reported on the application of diffusion heterogeneity to the classification of HNSCC lymph nodes.

The aim of this study was therefore to determine whether pretreatment nodal microstructural diffusion MRI characteristics can classify diseased nodes of patients with HNSCC from normal nodes of healthy volunteers.

2 | MATERIALS AND METHODS

2.1 | Patient population

Institutional review board approval was achieved and written informed consent obtained from all patients and healthy volunteers for this prospective study. From March 2010 to July 2017, 79 patients with histologically confirmed HNSCC with cervical nodal N2/N3 metastatic disease[14] were recruited prior to chemoradiotherapy or radiotherapy only. Five patients declined trial entry and two had neck dissection without any residual disease, resulting in 72 recruited patients. Exclusion criteria were related to MR imaging (i.e. claustrophobia, pacemaker, metallic implant, protocol deviations, poor image quality); 11 of 72 patients were excluded from analysis, two due to not attending the MRI scan and nine due to inappropriate MR imaging. Subsequently, data of 61 of 72 patients (mean age 57.7 years; range 25 to 79 years), 48 males (mean age 58.4 years; range 33 to 79 years) and 13 females (mean age 54.8 years; range 25 to 74 years), were analysed. Patient selection is illustrated in Figure 1. To compare the microstructural heterogeneity between diseased and healthy nodes, eight healthy volunteers were also recruited (mean age 43.4 years; range 19 to 65 years), three males (mean age 58 years; range 50 to 65 years) and five females (mean age 39.4 years; range 19 to 56 years), from March 2010 to March 2012. For the volunteers' scans, inclusion criteria were no previous malignancy and infective disease, and the exclusion criteria were contraindications to MR imaging.

2.2 | Nodal staging

Two experienced head and neck radiologists (T.B. and S.M., with 18 and 10 years of experience, respectively) reviewed all CT and anatomical MRI and performed neck ultrasound (US) in all patients, as per local standard of care. Cervical nodes were assessed as per the Union for International Cancer (UICC) TNM Classification of Malignant Tumour.[15] Per patient, the largest node within the neck was sampled by US-guided fine needle aspiration (FNA), and N2a (metastases in a single lymph node larger than 3 cm but not larger than 6 cm in greatest dimension), N2b (metastases in multiple ipsilateral nodes not larger than 6 cm) or N3 (metastases to one or more lymph nodes greater than 6 cm in greatest dimension) status was confirmed.

2.3 | Research MRI protocol

Patients pretreatment (2 to 4 weeks following the US FNA) and healthy volunteers were scanned on a 1.5-T MR scanner (MAGNETOM Avanto, Siemens AG, Erlargen, Germany) using the carotid coils in the supine position. Axial T2-weighted (Periodically Rotated Overlapping ParallEL Lines
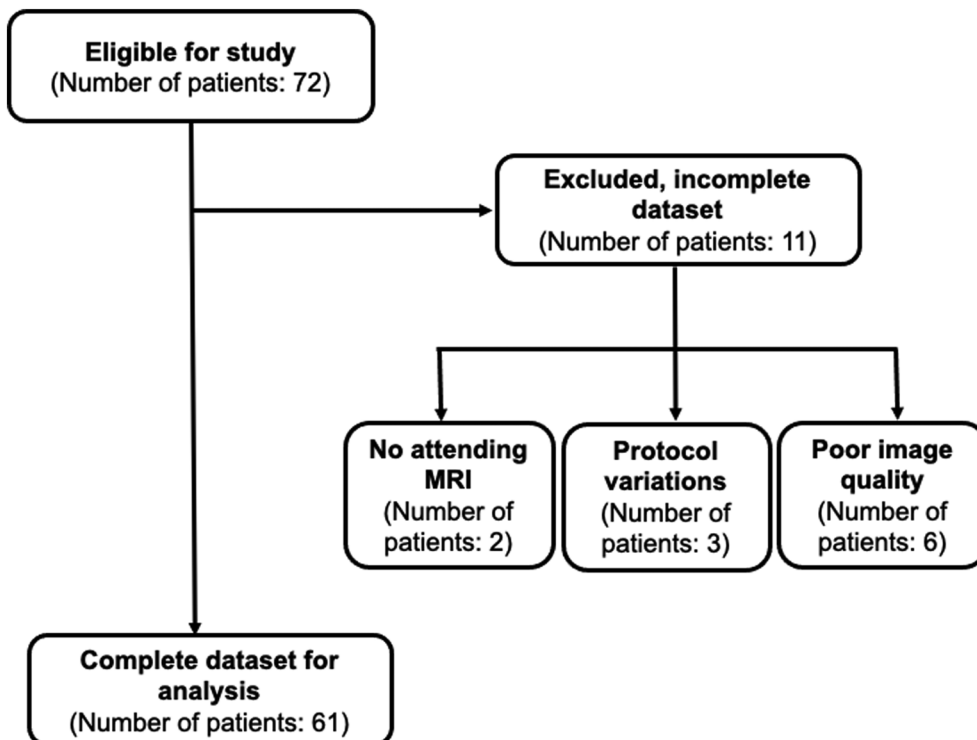
FIGURE 1 Flowchart illustrating the patient selection

with Enhanced Reconstruction [PROPELLER]) and DW images were acquired covering the neck, from the base of the skull to the upper thorax (Table 1). For the DW acquisition, images with six different b-values were acquired. Low b-value (b = 0, 50,100 s/mm^2) images were obtained for reflecting the signal loss mainly due to microcapillary perfusion, and the high b-value (b = 300, 600, 1000 s/mm^2) images for reflecting the signal loss, mainly of the diffusion component. This approach provided a clinically acceptable acquisition time of 4–5 min.

2.4 | Diffusion parameter extraction

Monoexponential (Equation 1) and stretched exponential (Equation 2) models were fitted to all pixels of the trace DW images of each slice using a least-squares fit for all b-values (MATLAB 2016; MathWorks Inc., Natick, MA, USA). A radiologist with 8 years of experience (H.S.), aware of the

**TABLE 1** T2- and diffusion-weighted pulse sequence parameters

Parameter	T2-weighted PROPELLER	Diffusion-weighted
Sequence	Turbo spin echo	Single-shot echo planar imaging
Slice orientation	Axial	Axial
Field of view (FOV) [mm (read) x mm (phase)]	180 × 180	206 × 206
Acquired (in-plane) matrix (read)	256	128
Reconstructed matrix (read)	256	256
Number of signal averages	1	4
Slice thickness (mm)	3	4
Slice gap (mm)	0.3	0.4
Number of slices	60	40 to 46
Parallel imaging reduction factor (GRAPPA)	2	2
Echo time (TE) (ms)	107	88
Repetition time (TR) (ms)	6,310	8700
Fat suppression	n/a	Short tau inversion recovery
Diffusion weightings (s/mm^2)	n/a	0, 50, 100, 300, 600 and 1000

PROPELLER, Periodically Rotated Overlapping ParallEL Lines with Enhanced Reconstruction.
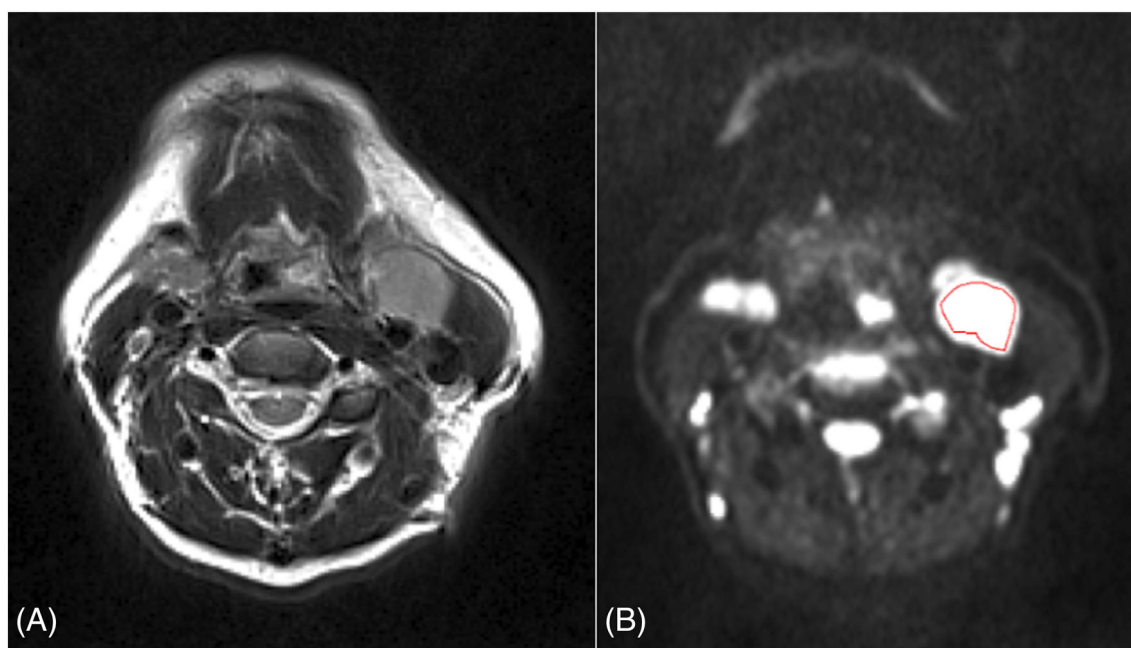
FIGURE 2 (A) Axial Periodically Rotated Overlapping ParallEL Lines with Enhanced Reconstruction (PROPELLER) T2-weighted image demonstrating diseased lymph nodes in a 52-year-old female patient. (B) Axial diffusion-weighted (DW) trace image (b = 300 s/mm^2) with the contoured regions of interest (ROI) in a 52-year-old female patient

location of the sampled nodes and nodal status at conventional staging, reviewed the precontrast T1-weighted, T2- weighted PROPELLER and trace DW images. Regions of interest (ROIs) were drawn using Jim 5.0 software (Xinapse Systems, Thorpe, Waterville, UK). For the patient groups, the radiologist contoured the ROIs on DW images of $b = 300$ s/mm^2 on each imaging slice containing the diseased node, encompassing all solid nodal tissue while avoiding any macroscopic necrotic and cystic areas (Figure 2A,B). For the healthy volunteers, the radiologist contoured one or more normal node ROIs per volunteer on DW images of $b = 300$ s/mm^2 on each imaging slice containing the specified node or nodes, taking care to only include nodal tissue, whereby each selected node was clearly visible and suitable to draw around (Figure 3A,B). For both cases, the ROIs from the DW images were electronically transferred to the corresponding calculated maps and a list of voxel-by-voxel diffusion parameter values were derived for each node (Figures 4A-C and 5A-C). For each patient and volunteer node, the histogram distribution of ADC, DDC and α values for the entire nodal volume were estimated (Figures 6A,B and 7A,B).
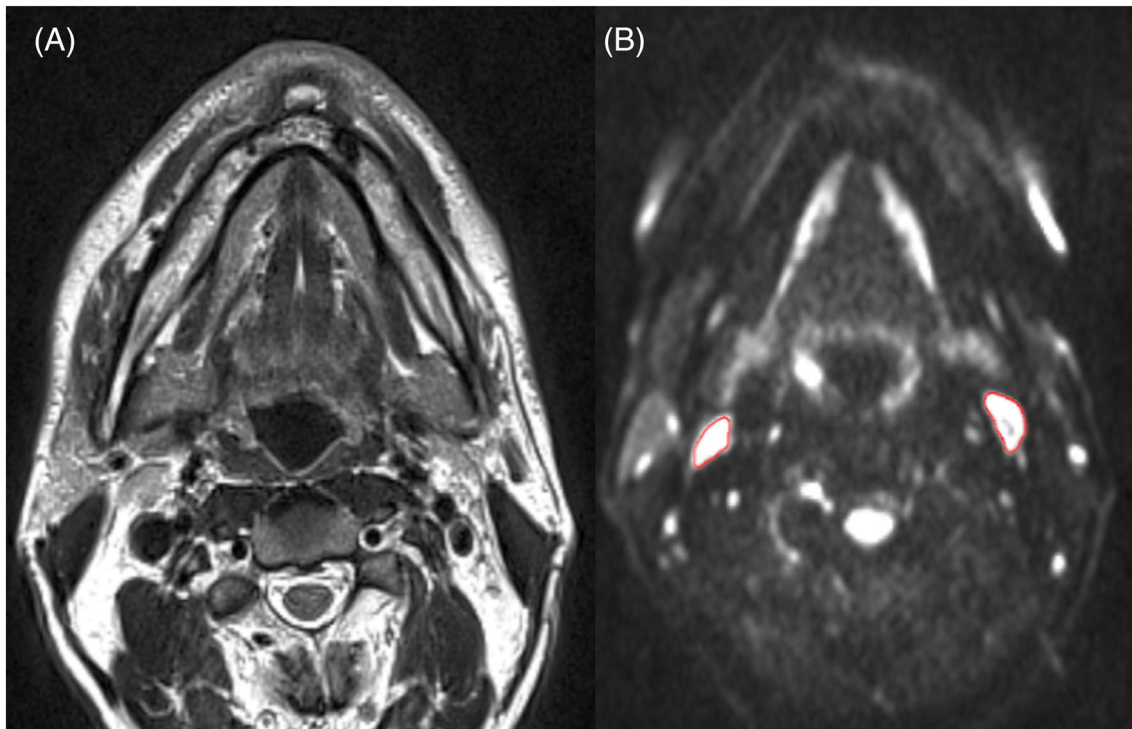
FIGURE 3 (A) Axial Periodically Rotated Overlapping ParallEL Lines with Enhanced Reconstruction (PROPELLER) T2-weighted image demonstrating healthy nodes in a 50-year-old healthy male volunteer. (B) Axial diffusion-weighted (DW) trace image ($b = 300$ s/mm^2) with the contoured regions of interest (ROI) in a 50-year-old healthy male volunteer
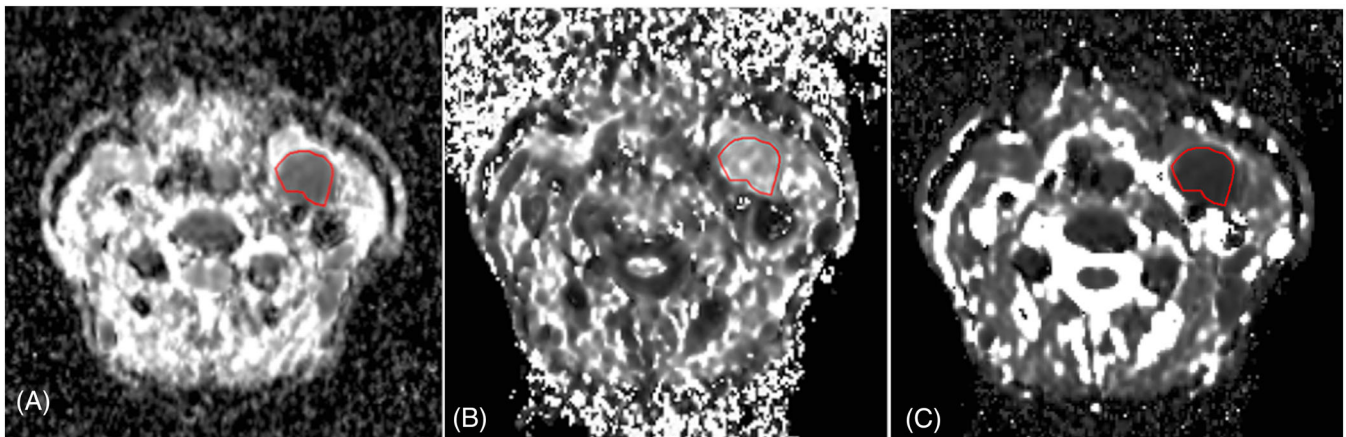
FIGURE 4 Calculated parametric diffusion maps: (A) apparent diffusion coefficient (ADC) map, (B) alpha (α) value map and (C) distributed diffusion coefficient (DDC) map with the contoured regions of interest (ROIs) from a 52-year-old female patient
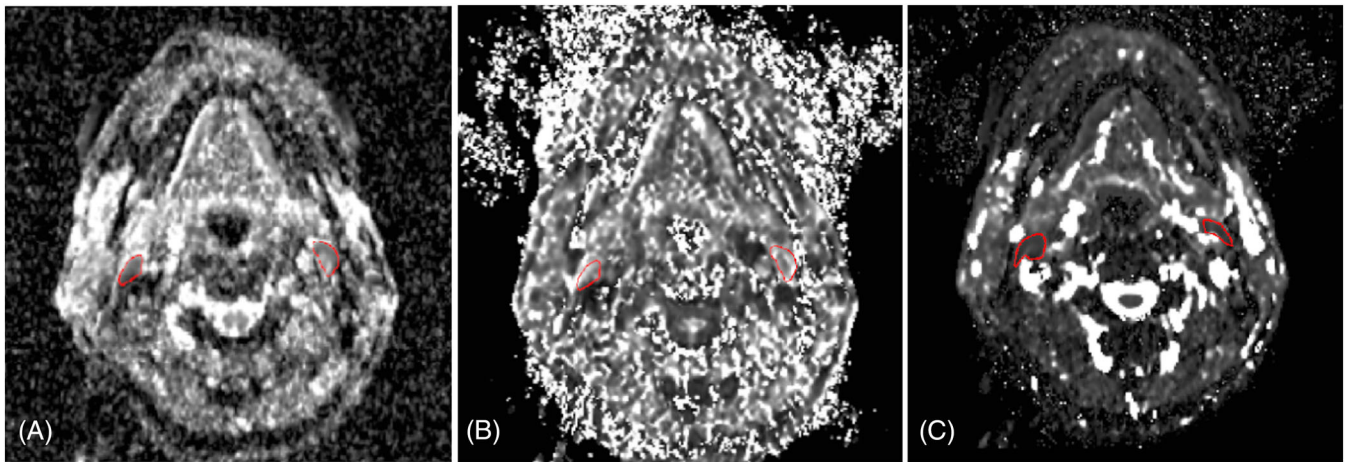
FIGURE 5 Calculated parametric diffusion maps: (A) apparent diffusion coefficient (ADC) map, (B) alpha (α) value map and (C) distributed diffusion coefficient (DDC) map with the contoured regions of interest (ROIs) from a 50-year-old healthy male volunteer
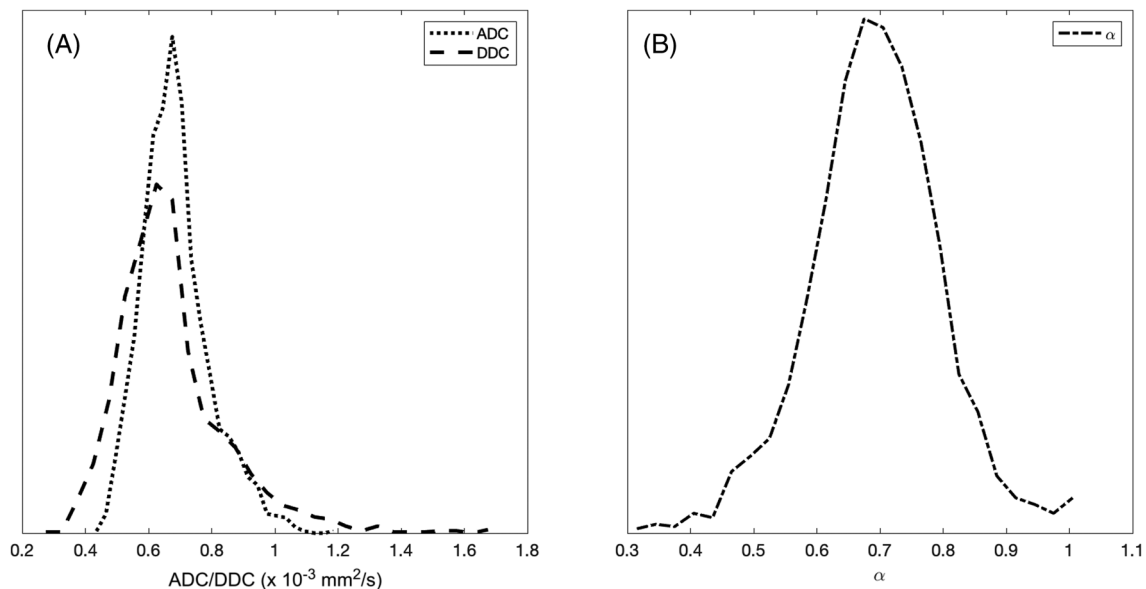
FIGURE 6 Histogram distributions of (A) apparent diffusion coefficient (ADC) and distributed diffusion coefficient (DDC) maps, and (B) alpha (α) value map from the entire contoured nodal volume of a 52-year-old female patient

2.5 | Statistical analysis

To assess if there were significant differences between the histogram metrics of ADC, DDC and α values of patients and healthy volunteers, each variable was compared between the two groups. For the statistical analysis, a Mann–Whitney test was utilised for the comparison between the two groups, taking into account that the data were not normally distributed. A significance level of 0.05 was used for all comparisons.

Receiver operating characteristic (ROC) analysis was performed to calculate the area under the ROC curve of any significantly different parameter. The sensitivity, specificity, accuracy, positive predictive value and negative predictive value of the cut-off values were also determined, using as threshold probability the maximum sensitivity and specificity. Statistical and ROC analysis were performed with Statistical Package for Social Sciences version 26.0 (SPPS Inc., Chicago, IL, USA).

2.6 | Interobserver reproducibility

To assess interobserver reproducibility, another radiologist with 6 years of experience (S.S.), aware of the location of the sampled nodes and nodal status at conventional staging, but blinded from the other radiologist's ROI drawings, reviewed the T2-weighted PROPELLER and trace DW
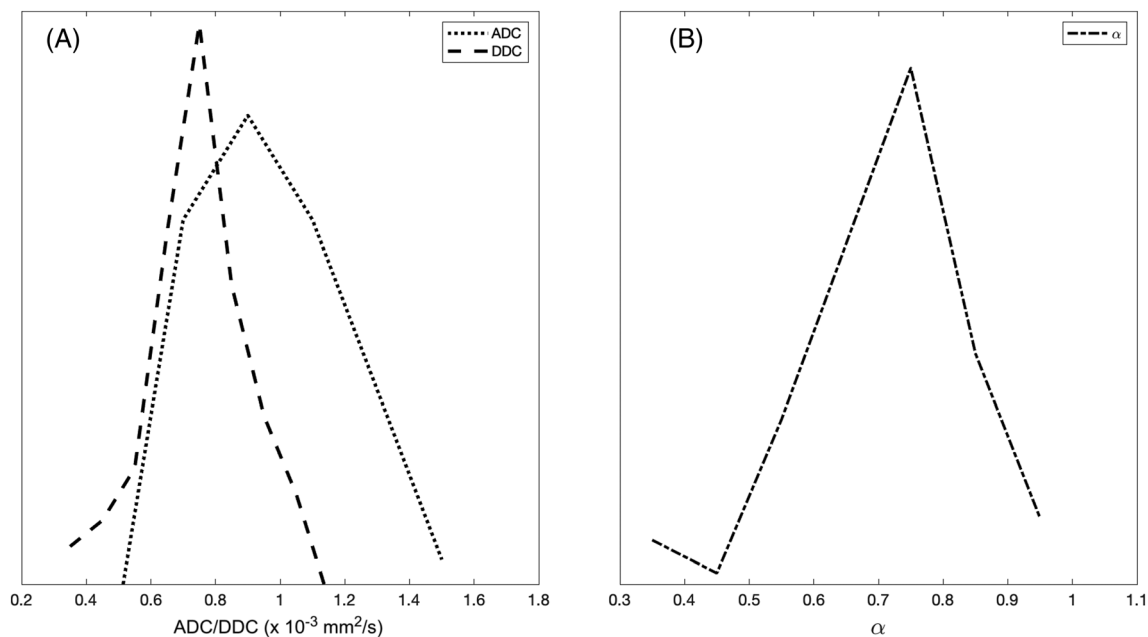
FIGURE 7 Histogram distributions of (A) apparent diffusion coefficient (ADC) and distributed diffusion coefficient (DDC) maps, and (B) alpha (α) value map from the entire contoured nodal volume of a 50-year-old healthy male volunteer

images. For 39 of 61 patients, the second radiologist contoured the ROIs on DW images of b = 300 s/mm^2 on each imaging slice containing the diseased node, encompassing all solid nodal tissue while avoiding any macroscopic necrotic and cystic areas, using Jim 5.0 software. Again, the ROIs from the DW images were electronically transferred to the corresponding calculated maps and the histogram distribution of the ADC, DDC and α values for the entire nodal volume for the 39 patients were estimated.

Interobserver agreement of ADC, DDC and α histogram metrics based on two different ROIs drawn by two radiologists were determined using intraclass correlation coefficients (ICCs). The ICCs were interpreted as follows: <0.40, poor reproducibility; 0.41–0.60, moderate reproducibility; 0.61–0.80, good reproducibility; and ≥0.81, excellent reproducibility.

2.7 | Logistic regression model

To define the histogram metrics of ADC, DDC and α values that significantly ($p < 0.05$) contributed to the nodal classification (diseased from healthy) and exhibited an excellent interobserver reproducibility, a score test was used. A logistic regression model was derived including the two histogram metrics with the higher scores and excellent interobserver reproducibility.[16] The ROC area under the curve was also calculated for the regression model. ROC and logistic regression analysis were performed with Statistical Package for Social Sciences version 26.0.

3 | RESULTS

In this study, 60 cancerous nodes (consisting of 86 to 11 493 voxels, median = 1663) were analysed from 61 patients, and 40 normal nodes (consisting of 37 to 523 voxels, median = 117) from the eight healthy volunteers. One lesion was excluded from the analysis, due to its cystic appearance in the DW images, not presenting any solid nodal tissue for the ROI drawing. Also, there was no precontrast T1-weighted high signal within the FNA sampled nodes to suggest haemorrhage-induced artifact. Patient demographics, age, gender, primary tumour site, TNM stage and lymph node size are listed in Table 2. Table 3 summarises the pretreatment histogram and statistical analysis of the ADC, DDC and α values, utilising the monoexponential and stretched exponential models of patients' and healthy volunteers' lymph nodes and their comparisons, respectively.

3.1 | Monoexponential model analysis

The 25th percentile, 75th percentile, maximum, median and mean nodal ADC values of the healthy volunteers were significantly lower compared with those of the patients (p ≤0.0001 to 0.0046; Table 3). The ADC nodal histogram distributions were more skewed (0.72 for patients and 0.16

TABLE 2 Patient and tumour characteristics

Patient no.	Gender	Age (y)	Primary tumour site of SCC	Primary tumour stage	Lymph node size (small axis [mm])
1	Male	67	Supraglottis and hypopharynx	T3 N2b M0	27
2	Female	55	Soft palate	T2 N2c M0	26
3	Female	74	Faucial tonsil	T1 N3 M0	7
4	Male	56	Right tonsil and piriform	T1 N2 M0	13
5	Male	56	Left faucial tonsil	T2 N2c M0	13
6	Male	67	Left inferior tonsil	T1 N2b M0	16
7	Male	56	Left BoT	T4 N2c M0	10
8	Male	63	Left faucial tonsil	T2 N2b M0	26
9	Male	60	Left pyriform fossa	T3 N2b M0	24
10	Male	48	Right pyriform fossa	T2 N2b M0	47
11	Female	52	Unknown	Tx N2 M0	16
12	Male	57	Right faucial tonsil	T2 N2b	30
13	Male	56	Right tonsil	T3 N2b M0	23
14	Male	57	Right tonsil	T2 N2b M0	17
15	Male	79	Right tonsil	T1 N2b M0	21
16	Male	52	Right glossotonsil sulcus	T4 N2c	16
17	Male	54	BoT and epiglottis	T3 N2c M0	20
18	Male	49	Right tonsil	pT4 N2b M0	20
19	Female	74	Left BoT	T2 N2b M0	32
20	Male	49	Left BoT and glossotonsillar sulcus	T3 N2b M0	14
21	Male	43	Right BoT	Tx N2c M0	14
22	Male	59	Left epiglottis	T2 N2c M1	24
23	Male	60	Right tonsil	T4 N2b M0	29
24	Female	47	Left tonsil	T2 N2b M0	18
25	Male	48	Right faucial tonsil	T3 N2b M0	21
26	Male	44	Right pyriform and aryepiglottic	T1 N2b Mx	28
27	Male	63	Right BoT	T4 N3 M0	46
28	Male	63	Right pyriform to oesophagus	T3 N2b M0	15
29	Female	49	Left pyriform fossa	T2 N2b M0	24
30	Male	67	Right tonsil	T4 N2c M0	15
31	Male	62	Right BoT	T4 N2c M0	22
32	Male	75	Right tonsil	T3 N2b M0	17
33	Female	25	Left tonsil	T2 N2b M0	17
34	Male	70	Left BoT	T4a N2b M0	21
35	Male	73	Right tonsil	T3 N2b M0	15
36	Male	64	Supraglottis	T3 N2c M0	18
37	Male	52	Right pyriform	T3 N2b M0	15
38	Male	54	Nasopharyngeal	T4 N3 M0	20
39	Male	76	Hypopharyngeal	T3 N2c M0	16
40	Male	68	Post oro/hypo-pharyngeal	T4a N2c M0	13
41	Male	67	Left tonsil	T3 N2c M0	22
42	Male	71	Left tonsil	T4b N2a M0	7
43	Male	62	Floor mouth/BoT	T2 N2a M0	10
44	Male	40	Left BoT	T2 N2c M0	27
45	Male	58	Right tonsil	T2 N2b M0	21
46	Male	57	Supraglottis	T3N2cM0	16

TABLE 2 (Continued)

Patient no.	Gender	Age (y)	Primary tumour site of SCC	Primary tumour stage	Lymph node size (small axis [mm])
47	Male	57	Hypopharynx	T2 N2c M0	25
48	Female	70	Unknown	T2 N2b M0	15
49	Male	59	Right tonsil	T4a N2b M0	16
50	Male	33	Right tonsil	T3 N2b Mo	15
51	Female	61	BoT	T2 N2b Mo	27
52	Male	41	Right tonsil	pT2a N2a M0	27
53	Male	59	Oropharynx (BoT)	T3 N2c M0	25
54	Male	52	Unknown	T3 N2b M0	21
55	Female	46	Left tonsil	T2 N2a M0	22
56	Female	61	Left tonsil	T2 N2b M0	31
57	Male	53	Post pharyngeal	T4a N2c M0	13
58	Male	68	Oropharyngeal down from GTS	T3 N2c M0	23
59	Female	56	Left tonsil	T2 N2b M0	21
60	Female	43	Nasopharyngeal	T1 N2 M0	12
61	Male	58	Oropharyngeal L BoT/GTS	T1 N2b M0	25

N2a: single >3 cm but ≤6 cm in greatest dimension.
N2b: metastases in multiple ipsilateral nodes <6 cm.
N3: node >6 cm.
Abbreviations: BoT, base of tongue; GTS, glossotonsillar sulcus; N/A, not available; SCC, squamous cell carcinoma.

for healthy volunteers) and more 'peaked' (kurtotic) (3.94 and 3.05 for patients and healthy volunteers, respectively) in patients compared with healthy volunteers ($p \leq 0.001$ for both skewness and kurtosis). Table 4 summarises the diagnostic performance of the ADC histogram metrics in the classification of diseased from healthy nodes. Maximum ADC values exhibited the highest area under the curve ([AUC] 0.892), sensitivity (86.2%) and specificity (77.5%) among all the histogram metrics, followed by skewness and kurtosis (AUC = 0.765, sensitivity = 78.3% and specificity = 75% for skewness, and AUC = 0.746, sensitivity = 58.3% and specificity = 77.5% for kurtosis).

3.2 | Stretched exponential model analysis

The 25th percentile, median and mean nodal α values of healthy volunteers were significantly lower compared with those of patients ($p \leq 0.0001$ to 0.0044; Table 3). The α nodal histogram distributions of patients were more kurtotic (2.817 for patients and 2.20 for healthy volunteers; $p = 0.0013$). For both groups, the α nodal histogram distributions were negatively skewed, exhibiting a longer tail in patients than in healthy volunteers, but without reaching statistical significance ($p = 0.1279$).

The 25th percentile, maximum, median and mean nodal DDC values of the healthy volunteers were lower compared with those of the patients (Table 3), reaching statistical significance ($p \leq 0.0001$ to 0.0114). The DDC nodal histogram distributions demonstrated significant increased skewness (1.49 and 1.03 for patients and healthy volunteers, respectively) and kurtosis (6.82 for patients and 4.20 for healthy volunteers) in patients compared with healthy volunteers ($p = 0.0114$ for skewness and $p = 0.0021$ for kurtosis).

Table 4 summarises the diagnostic performance of the DDC and α value histogram metrics in the classification of diseased from healthy nodes. From the DDC nodal histogram analysis, maximum DDC exhibited the highest AUC (0.772), sensitivity (83.3%) and specificity (62.5%). From the α value histogram analysis, the 25th percentile presented the highest AUC (0.761), sensitivity (80.0%) and specificity (67.5%).

3.3 | Interobserver variability

The ICC results for each one of the MR diffusion histogram metrics are presented in Table 5. From the ADC histogram analysis, interobserver agreement was excellent for the minimum (0.84), 25th percentile (0.94), 75th percentile (0.85), median (0.92) and mean (0.88), and poor for maximum (0.37), skewness (0.19) and kurtosis (0.25). From the α value histogram analysis, interobserver agreement was excellent for the 25th percentile (0.78), 75th percentile (0.77), median (0.80), mean (0.79) and skewness (0.80), and fair for minimum (0.58) and kurtosis (0.60). On the other

TABLE 3 Apparent diffusion coefficient (ADC), distributed diffusion coefficient (DDC) and alpha (α) value histogram results from patients and healthy volunteers (including mean ± standard deviation for all the metrics, and median and interquartile range for kurtosis), and p values derived from the statistical comparisons

	Patients (60 patients, 60 nodes)	Healthy volunteers (8 volunteers, 40 nodes)	p value (between patients and healthy volunteers)
ADC (10^{-3} mm^2/s)			
Minimum	0.43 ± 0.19	0.46 ± 0.17	0.4543
25th percentile	0.77 ± 0.14	0.67 ± 0.12	0.0004[a]
75th percentile	1.00 ± 0.23	0.88 ± 0.15	0.0046[a]
Maximum	1.66 ± 0.42	1.14 ± 0.23	<0.0001[a]
Median	0.88 ± 0.17	0.78 ± 0.13	0.0046[a]
Mean	0.90 ± 0.17	0.78 ± 0.13	0.0010[a]
Skewness	0.72 ± 0.61	0.16 ± 0.51	<0.0001[a]
Kurtosis	3.94 [2.99, 5.32]	3.05 [2.35, 3.52]	<0.0001[a]
DDC (10^{-3} mm^2/s)			
Minimum	0.34 ± 0.21	0.36 ± 0.20	0.4563
25th percentile	0.80 ± 0.24	0.68 ± 0.16	0.0029[a]
75th percentile	1.12 ± 0.44	0.97 ± 0.24	0.0516
Maximum	2.70 ± 0.95	1.85 ± 0.92	<0.0001[a]
Median	0.96 ± 0.44	0.81 ± 0.17	0.0097[a]
Mean	0.99 ± 0.34	0.84 ± 0.19	0.0121[a]
Skewness	1.49 ± 1.02	1.03 ± 1.28	0.0114[a]
Kurtosis	6.82 [4.95,11.65]	4.20 [2.60,9.61]	0.0021[a]
α			
Minimum	0.29 ± 0.13	0.34 ± 0.14	0.1294
25th percentile	0.67 ± 0.08	0.60 ± 0.09	<0.0001[a]
75th percentile	0.88 ± 0.08	0.89 ± 0.10	0.7721
Maximum	1.00 ± 0.0	0.99 ± 0.03	0.0611
Median	0.78 ± 0.08	0.73 ± 0.08	0.0047[a]
Mean	0.77 ± 0.06	0.73 ± 0.07	0.0044[a]
Skewness	−0.318 ± 0.444	−0.13 ± 0.52	0.1279
Kurtosis	2.817 [2.42,3.23]	2.20 [1.86, 297]	0.0013[a]

[a]statistically significant difference.

hand, from the DDC histogram analysis, interobserver agreement was excellent for the minimum (0.89), 25th percentile (0.826), median (0.92) and mean (0.85), fair to good for the 75th percentile (0.62), and poor for maximum (0.22), kurtosis (0.10) and skewness (0.19).

3.4 | Calculation of the logistic regression model

The score results from the MR diffusion histogram metrics are listed in Table 6. For the calculation of the logistic regression model, the two highest score metrics with excellent interobserver reproducibility, which significantly contributed to the model, were determined. These metrics were the mean ADC (score test = 12.332) and the 25th percentile α (score test = 12.800), presenting a statistically significant contribution to the model (p ≤0.001 for both metrics). Subsequently, the logistic regression equation for the classification of diseased from healthy nodes is:

$$\ln(Odds) = -9.5 + 5.037 \times (\text{Mean}_{ADC}) + 8.887 \times (25\text{th Percentile}_\alpha). \tag{3}$$

ROC analysis was performed to determine the best cut-off as nodal classifier using the regression model (Figure 8). Based on this, the optimal cut-off value was 0.1604, the AUC 0.80, sensitivity 82%, specificity 68%, positive predictive value (PPV) 79% and negative predictive value (NPV) 71%, and the overall accuracy was 76%.

TABLE 4 Receiver operating characteristic (ROC) analysis (area under the curve [AUC], cut-off values, sensitivity, specificity, positive predictive value, negative predictive value) for the apparent diffusion coefficient (ADC), distributed diffusion coefficient (DDC) and alpha (α) values histogram metrics

	AUC	Cut-off value	Sensitivity (%)	Specificity (%)	Positive predictive value	Negative predictive value
ADC (10^{-3} mm²/s)						
25th percentile	0.708	0.699	68.3	67.5	75.9	58.7
75th percentile	0.666	0.913	63.3	65.0	73.1	54.2
Maximum	0.892	1.278	86.2	77.5	84.7	79.5
Median	0.666	0.818	60.0	67.5	73.5	52.9
Mean	0.692	0.825	61.7	67.5	74.0	54.0
Skewness	0.765	0.371	78.3	75.0	82.5	69.8
Kurtosis	0.746	3.551	58.3	77.5	79.5	74.6
DDC (10^{-3} mm²/s)						
25th percentile	0.675	0.730	63.3	70.0	76.0	56.0
Maximum	0.772	1.734	83.3	62.5	76.9	71.4
Median	0.652	0.792	68.3	52.5	68.3	52.5
Mean	0.648	0.893	56.7	60.0	68.0	48.0
Skewness	0.649	0.848	76.7	52.5	70.7	60.0
Kurtosis	0.680	5.010	73.3	60.0	73.3	60.0
α						
25th percentile	0.761	0.639	80.0	67.5	78.7	69.2
Median	0.666	0.777	55.0	72.5	75.0	52.8
Mean	0.652	0.792	68.3	52.5	68.3	52.5
Kurtosis	0.688	2.361	76.7	57.5	73.0	62.2

4 | DISCUSSION

The current study aimed to determine whether pretreatment microstructural diffusion MRI characteristics could classify diseased nodes of patients with HNSCC from normal nodes of healthy volunteers. We utilised two exponential models for the histogram analysis, namely, the well-established monoexponential model for the ADC assessment, and the stretched exponential model for the assessment of DDC and α values; the latter allows determination of intravoxel diffusion heterogeneity.[17,18] Histogram analysis was utilised for ADC and DDC assessment to provide a measure of intervoxel heterogeneity.[9,12,19] We focused on the lymph nodes, because the status of the cervical lymph nodes is considered to be the most important prognostic factor.[4]

Our first finding was that the 25th percentile, 75th percentile, maximum, median and mean nodal ADC values were significantly higher in patients than in healthy volunteers. These higher ADC values denote that the averaged water diffusivity of the diseased nodes is higher compared with that of the healthy nodes. This can be ascribed to the presence of micronecrosis (not apparent from the DW images) in the diseased nodes, which is averaged with the cellular components, resulting in higher median and mean ADC (ADC estimates) in patients than in healthy volunteers. The presence of micronecrosis is also related to the increased maximum ADC values in patients, presenting the highest diagnostic performance across all the MR diffusion histogram metrics. Other researchers have reported significant differences between benign and metastatic cervical lymph nodes. Sumi et al.[20] reported higher mean ADC values in metastatic lymph nodes than in benign nodes, highlighting that nodal necrosis is one of the characteristics of malignant nodes, whereas Vandecaveye et al.[21] and de Bondt et al.[5] reported higher ADC values in benign compared with metastatic nodes. However, in these studies,[5,20–22] the patient population was smaller (ranging from 16 to 33 patients) than in the current study (60 patients), no lymph nodes from healthy volunteers were included, and a mean ADC value was used for the ADC comparison. To our knowledge, there are no other studies utilising comparison of ADC histogram analysis of lymph nodes in patients with HNSCC and healthy volunteers.

Second, the ADC histogram analysis showed that nodal disease could be identified by assessing intervoxel diffusion heterogeneity. The significantly increased kurtosis in diseased patient nodes suggests a decrease in heterogeneity for the larger part of the diseased nodal volume with voxels with a narrower range of ADC values compared with the broader distribution of healthy volunteer nodes. The increased skewness in patients resulted from distributions with a longer right tail, which may be due to micronecrosis, as suggested by Scalco et al.,[23] resulting in higher ADC values and in asymmetric ADC distributions. In HNSCC primary tumours, the value of ADC histogram analysis has been applied to assess

TABLE 5 Interobserver variability of apparent diffusion coefficient (ADC), distributed diffusion coefficient (DDC) and alpha (α) values histogram metrics

MR diffusion histogram metric	ICC
ADC (10^{-3} mm^2/s)	
Minimum	0.84
25th percentile	0.94
75th percentile	0.85
Maximum	0.37
Median	0.92
Mean	0.88
Skewness	0.19
Kurtosis	0.25
DDC (10^{-3} mm^2/s)	
Minimum	0.89
25th percentile	0.95
75th percentile	0.62
Maximum	0.22
Median	0.92
Mean	0.85
Skewness	0.19
Kurtosis	0.10
α	
Minimum	0.58
25th percentile	0.78
75th percentile	0.77
Maximum	N/A
Median	0.80
Mean	0.79
Skewness	0.80
Kurtosis	0.60

Abbreviation: ICC, interclass correlation coefficient.

TABLE 6 Score tests of apparent diffusion coefficient (ADC), distributed diffusion coefficient (DDC) and alpha (α) value histogram metrics for the calculation of the logistic regression model with p values denoting the statistically significant contribution to the model

MR diffusion histogram metric	Score test for classification, p value significant contribution to the model
ADC (10^{-3} mm^2/s)	
25th percentile	11.439 ($p = 0.001$)
75th percentile	8.527 ($p = 0.003$)
Median	9.464 ($p = 0.002$)
Mean	12.332 ($p \leq 0.001$)
DDC (10^{-3} mm^2/s)	
25th percentile	7.678 ($p = 0.006$)
Median	4.017 ($p = 0.45$)
Mean	5.811 ($p = 0.016$)
α	
25th percentile	12.800 ($p \leq 0.001$)
Median	6.633 ($p = 0.010$)
Mean	7.014 ($p = 0.008$)
Kurtosis	7.182 ($p = 0.007$)
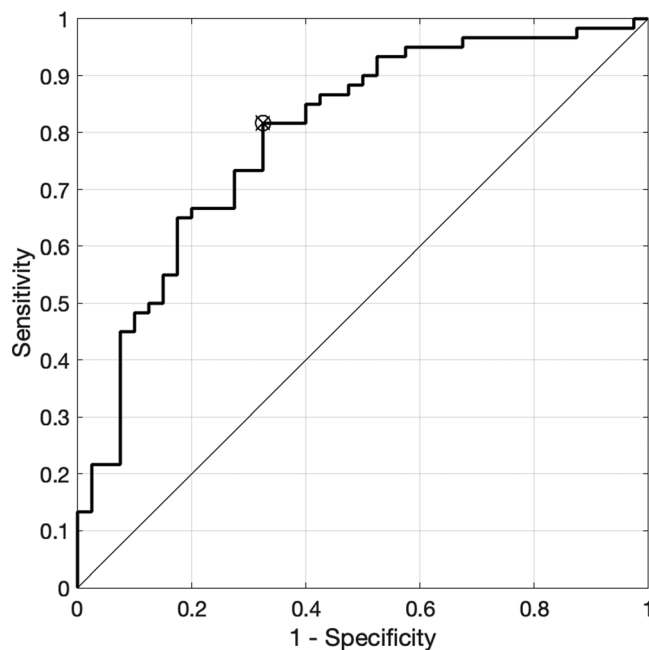
FIGURE 8 Receiver operating characteristic (ROC) analysis of the logistic regression model to classify diseased from healthy nodes, presenting the best cut-off value as nodal classifier

differentiation of tumour grades[13] and reflect tumour heterogeneity.[19] In our study, the high diagnostic performance of ADC skewness and kurtosis suggests that a similar approach is valid for classification of lymph node disease in HNSCC patients from the lymph nodes of healthy volunteers.

The assessment of the intravoxel (α value) and the intervoxel (DDC histogram distribution) MRI diffusion heterogeneity, utilising the stretched exponential model, could also help to classify diseased nodes from those of healthy volunteers. The nodal DDC histogram analysis revealed more 'peaked' and skewed distributions with significantly higher 25th percentile, maximum, median and mean DDC values in patients than in healthy volunteers, similar to the ADC histogram analysis. The statistical comparisons revealed significantly higher median, mean, 25th and 75th percentile and kurtosis nodal α values in patients than in healthy volunteers, recommending the 25th percentile as a reliable nodal disease classifier. The α value is a stretching parameter taking into account the distribution of the diffusion coefficients within a voxel and defining the deviation of the signal attenuation from the monoexponential decay[8]; as the α value approaches 1, the signal attenuation approaches monoexponential behaviour. A higher α value in patients supports lower intravoxel homogeneity in diffusion coefficients, reflecting less heterogeneity caused by intravascular perfusion, and extravascular diffusion processes existing within the same voxel. Absolute ADC values for disease states and healthy tissues reflect a single predominant tissue feature (e.g. cellularity), but they are also influenced by other tissue characteristics (e.g. vascularity). Therefore, bi-exponential models have been introduced, including low diffusion and fast diffusion (perfusion) components, and applied to nodal disease reporting the predictor role of f value.[24,25] However, the IVIM analysis assumes only perfusion and diffusion components without accounting either for the slow exchange between them or for other diffusion rates. The stretched exponential model overcomes these assumptions and the stretching parameter has been reported to be more sensitive to the microstructural properties of cervical tumours than the ADC,[26] and is able to differentiate high grade gliomas from normal tissues,[18] as well as discriminate low to high grade gliomas[27] and different tumour stages of naso-pharyngeal carcinoma,[28] while being stable to noise.[8]

The assessment of the interobserver reproducibility showed that the parameters related to the shape of ADC and DDC distributions of each node, such as skewness and kurtosis, varied between different observers. Whereas distribution metrics unrelated to the shape of the distribution, such as the 25th percentile, mean and median, exhibited excellent interobserver agreement. Ren et al.[29] have reported that the ROI selection influenced the ADC histogram analysis between two observers. Liu et al.[30] also reported excellent agreement between two observers in the ADC histogram results, apart from skewness and kurtosis. In ADC and DDC histogram analysis, skewness and kurtosis are metrics assessing the intervoxel heterogeneity across the nodal volume, although they are dependent on each observer. Deviations in the drawn edges of the ROI nodes differently assessed by the observers can conclude in different results. Therefore, a reliable nodal disease classifier should not only present higher diagnostic performance in one observer, but also be reproducible between observers.

The logistic regression analysis revealed that the best two classifiers for the nodal disease in the linear regression model were the mean ADC and 25th percentile α value. In detail, mean ADC expresses a regional estimate of the water diffusivity across the whole volume of the node; and

the 25th percentile α value represents an intravoxel heterogeneity estimate. In the literature, several validation models have been presented as disease classifiers in prostate cancer utilising multiparametric MRI,[31] in breast cancer using multimodality imaging (PET and mpMRI),[32] and in thyroid nodules with bi-parametric MRI.[33] Our model not only includes a regional averaged estimate of the water diffusivity, but also takes into account the effect of heterogeneity within each voxel, resulting in an integrated classifier for nodal disease using only DW-MRI.

Our study has some limitations. DW-MRI in the head and neck is challenging due to the introduced magnetic field inhomogeneities across this area, resulting in DW images with low signal intensity and distortions.[34] In this study, DW images of six b-values were acquired, to ensure sufficient signal intensity covering the multiple exponential effects and resulting in a clinically acceptable scan time of 4–5 min. More accurate quantitation may be possible using a larger number of b-values,[35] but would limit clinical acceptability. The normal lymph nodes sample size was not exactly equal to the diseased lymph nodes sample size, but the fact that our data were significant suggests they are valid. Moreover, there is a lack of pathology confirmation of the healthy lymph nodes, although no previous malignancy and infective disease were necessary for the healthy volunteers' participation. Furthermore, following the institutional healthy volunteer scanning regulations, all the healthy volunteers' scans were reviewed by a radiologist to confirm that they were healthy before their inclusion in the study.

In summary, our quantitative analysis suggests that the nodal microstructural diffusion MRI characteristics assessed by ADC, DDC and α value histogram analysis can classify diseased from normal lymph nodes. Our linear regression model, taking into account the regional MR diffusion across the whole nodal volume (mean ADC) and intravoxel (25th percentile α value) MR diffusion heterogeneity, demonstrated high diagnostic performance as a nodal disease classifier. Future studies are required to further investigate the utility of the linear regression model, including a regional diffusion estimate and intravoxel/intervoxel diffusion heterogeneity as a predictor of treatment response.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

REFERENCES

1. Kim ES, Yoon DY, Moon JY, Baek S, Han YM, Seo YL. Detection of loco-regional recurrence in malignant head and neck tumors: a comparison of CT, MRI, and FDG PET-CT. *Acta Radiol*. 2019;60(2):186-195.
2. Padhani AR, Liu G, Mu-Koh D, et al. Diffusion-weighted magnetic resonance imaging as a cancer biomarker: consensus and recommendations. *Neoplasia*. 2009;11(2):102-125.
3. Koh DM, Collins DJ. Diffusion-weighted MRI in the body: applications and challenges in oncology. *Am J Roentgenol*. 2007;188(6):1622-1635.
4. Driessen JP, van Kempen PM, van der Heijden GJ, et al. Diffusion-weighted imaging in head and neck squamous cell carcinomas: a systematic review. *Head Neck*. 2015;37(3):440-448.
5. de Bondt RB, Hoeberigs MC, Nelemans PJ, et al. Diagnostic accuracy and additional value of diffusion-weighted imaging for discrimination of malignant cervical lymph nodes in head and neck squamous cell carcinoma. *Neuroradiology*. 2009;51(3):183-192.
6. Zhou M, Lu B, Lv G, et al. Differential diagnosis between metastatic and non-metastatic lymph nodes using DW-MRI: a meta-analysis of diagnostic accuracy studies. *J Cancer Res Clin Oncol*. 2015;141(6):1119-1130.
7. Le Bihan D, Breton E, Lallemand D, Aubin ML, Vignaud J, Laval-Jeantet M. Separation of diffusion and perfusion in intravoxel incoherent motion MR imaging. *Radiology*. 1988;168(2):497-505.
8. Bennett KM, Hyde JS, Schmainda KM. Water diffusion heterogeneity index in the human brain is insensitive to the orientation of applied magnetic field gradients. *Magn Reson Med*. 2006;56(2):235-239.
9. Kwee TC, Galbán CJ, Tsien C, et al. Comparison of apparent diffusion coefficients and distributed diffusion coefficients in high-grade gliomas. *J Magn Reson Imaging*. 2010;31(3):531-537.
10. Burrell RA, McGranahan N, Bartek J, Swanton C. The causes and consequences of genetic heterogeneity in cancer evolution. *Nature*. 2013;501(7467): 338-345.
11. Ganeshan B, Miles KA. Quantifying tumour heterogeneity with CT. *Cancer Imaging*. 2013;13:140-149.
12. Tozer DJ, Jager HR, Danchaivijitr N, et al. Apparent diffusion coefficient histograms may predict low-grade glioma subtype. *NMR Biomed*. 2007;20(1): 49-57.
13. Ahn SJ, Choi SH, Kim YJ, et al. Histogram analysis of apparent diffusion coefficient map of standard and high B-value diffusion MR imaging in head and neck squamous cell carcinoma: a correlation study with histological grade. *Acad Radiol*. 2012;19(10):1233-1240.
14. Edge SB, Compton CC. The American Joint Committee on Cancer: the 7th edition of the AJCC cancer staging manual and the future of TNM. *Ann Surg Oncol*. 2010;17(6):1471-1474.
15. Sobin LH. *TNM classification of malignant tumours*. 7th ed. Chichester, UK and Hoboken, NJ: John Wiley and Sons; 2009:255-287.
16. Menard S. An introduction to logistic regression diagnostics. *Appl Logistic Reg Anal*. 1995;106:58-79.

17. Bennett KM, Schmainda KM, Bennett RT, Rowe DB, Lu H, Hyde JS. Characterization of continuously distributed cortical water diffusion rates with a stretched-exponential model. *Magn Reson Med*. 2003;50(4):727-734.

18. Kwee T, Galban C, Ivancevic M, et al. Intravoxel water diffusion heterogeneity imaging of human high-grade gliomas. *NMR Biomed*. 2010;23(2): 179-187.

19. de Perrot T, Lenoir V, Domingo Ayllon M, Dulguerov N, Pusztaszeri M, Becker M. Apparent diffusion coefficient histograms of human papillomavirus-positive and human papillomavirus-negative head and neck squamous cell carcinoma: assessment of tumor heterogeneity and comparison with histo-pathology. *Am J Neuroradiol*. 2017;38(11):2153-2160.

20. Sumi M, Sakihama N, Sumi T, et al. Discrimination of metastatic cervical lymph nodes with diffusion-weighted MR imaging in patients with head and neck cancer. *Am J Neuroradiol*. 2003;24(8):1627-1634.

21. Vandecaveye V, De Keyzer F, Vander Poorten V, et al. Head and neck squamous cell carcinoma: value of diffusion-weighted MR imaging for nodal staging. *Radiology*. 2009;251(1):134-146.

22. Perrone A, Guerrisi P, Izzo L, et al. Diffusion-weighted MRI in cervical lymph nodes: differentiation between benign and malignant lesions. *Eur J Radiol*. 2011;77(2):281-286.

23. Scalco E, Marzi S, Sanguineti G, Vidiri A, Rizzo G. Characterization of cervical lymph-nodes using a multi-parametric and multi-modal approach for an early prediction of tumor response to chemo-radiotherapy. *Phys Med*. 2016;32(12):1672-1680.

24. Hauser T, Essig M, Jensen A, et al. Characterization and therapy monitoring of head and neck carcinomas using diffusion-imaging-based intravoxel incoherent motion parameters-preliminary results. *Neuroradiology*. 2013;55(5):527-536.

25. Hauser T, Essig M, Jensen A, et al. Prediction of treatment response in head and neck carcinomas using IVIM-DWI: Evaluation of lymph node metasta-sis. *Eur J Radiol*. 2014;83(5):783-787.

26. Winfield JM, Orton MR, Collins DJ, et al. Separation of type and grade in cervical tumours using non-mono-exponential models of diffusion-weighted MRI. *Eur Radiol*. 2017;27(2):627-636.

27. Bai Y, Lin Y, Tian J, et al. Grading of gliomas by using monoexponential, biexponential, and stretched exponential diffusion-weighted MR imaging and diffusion kurtosis MR imaging. *Radiology*. 2015;278(2):496-504.

28. Lai V, Lee VH, Lam KO, Sze HC, Chan Q, Khong PL. Intravoxel water diffusion heterogeneity MR imaging of nasopharyngeal carcinoma using stretched exponential diffusion model. *Eur Radiol*. 2015;25(6):1708-1713.

29. Ren JL, Yuan Y, Li XX, Shi YQ, Tao XF. Histogram analysis of apparent diffusion coefficient maps in the prognosis of patients with locally advanced head and neck squamous cell carcinoma: Comparison of different region of interest selection methods. *Eur J Radiol*. 2018;106:7-13.

30. Liu S, Zhang Y, Chen L, et al. Whole-lesion apparent diffusion coefficient histogram analysis: significance in T and N staging of gastric cancers. *BMC Cancer*. 2017;17(1):1-9.

31. Dikaios N, Giganti F, Sidhu HS, et al. Multi-parametric MRI zone-specific diagnostic model performance compared with experienced radiologists for detection of prostate cancer. *Eur Radiol*. 2019;29(8):4150-4159.

32. Leithner D, Horvat JV, Bernard-Davila B, et al. A multiparametric [(18)F]FDG PET/MRI diagnostic model including imaging biomarkers of the tumor and contralateral healthy breast tissue aids breast cancer diagnosis. *Eur J Nucl Med Mol Imaging*. 2019;46(9):1878-1888.

33. Wang H, Wei R, Liu W, Chen Y, Song B. Diagnostic efficacy of multiple MRI parameters in differentiating benign vs. malignant thyroid nodules. *BMC Med Imaging*. 2018;18(1):1-9.

34. Thoeny HC, De Keyzer F, King AD. Diffusion-weighted MR imaging in the head and neck. *Radiology*. 2012;263(1):19-32.

35. Fujima N, Sakashita T, Homma A, et al. Advanced diffusion models in head and neck squamous cell carcinoma patients: Goodness of fit, relationships among diffusion parameters and compariosn with dynamic contrast-enhanced perfusion. *Magn Reson Med*. 2016;36:16-23.