











## RESEARCH ARTICLE

# Metabolite selection for machine learning in childhood brain tumour classification

Dadi Zhao<sup>1,2</sup>  | James T. Grist<sup>1,2</sup>  | Heather E.L. Rose<sup>1,2</sup>  |  
 Nigel P. Davies<sup>1,2,3</sup>  | Martin Wilson<sup>4</sup>  | Lesley MacPherson<sup>5</sup> |  
 Laurence J. Abernethy<sup>6</sup> | Shivaram Avula<sup>6</sup>  | Barry Pizer<sup>7</sup> |  
 Daniel R. Gutierrez<sup>8,9</sup> | Tim Jaspán<sup>8,10</sup>  | Paul S. Morgan<sup>8,9,11</sup>  |  
 Dipayan Mitra<sup>12</sup> | Simon Bailey<sup>13</sup>  | Vijay Sawlani<sup>14,15</sup>  |  
 Theodoros N. Arvanitis<sup>1,2,16</sup>  | Yu Sun<sup>1,2,17</sup>  | Andrew C. Peet<sup>1,2</sup> 

<sup>1</sup>Institute of Cancer and Genomic Sciences, University of Birmingham, Birmingham, UK

<sup>2</sup>Oncology, Birmingham Children's Hospital, Birmingham, UK

<sup>3</sup>Imaging and Medical Physics, University Hospitals Birmingham, Birmingham, UK

<sup>4</sup>Centre for Human Brain Health, University of Birmingham, Birmingham, UK

<sup>5</sup>Radiology, Birmingham Children's Hospital, Birmingham, UK

<sup>6</sup>Radiology, Alder Hey Children's Hospital, Liverpool, UK

<sup>7</sup>Paediatric Oncology, Alder Hey Children's Hospital, Liverpool, UK

<sup>8</sup>Children's Brain Tumour Research Centre, University of Nottingham, Nottingham, UK

<sup>9</sup>Medical Physics, Nottingham University Hospitals NHS Trust, Nottingham, UK

<sup>10</sup>Neuroradiology, Nottingham University Hospitals NHS Trust, Nottingham, UK

<sup>11</sup>Division of Clinical Neuroscience, University of Nottingham, Nottingham, UK

<sup>12</sup>Neuroradiology, The Newcastle upon Tyne Hospitals, Newcastle upon Tyne, UK

<sup>13</sup>Paediatric Oncology, Great North Children's Hospital, Newcastle upon Tyne, UK

<sup>14</sup>Radiology, Queen Elizabeth Hospital Birmingham, Birmingham, UK

<sup>15</sup>School of Psychology, University of Birmingham, Birmingham, UK

<sup>16</sup>Institute of Digital Healthcare, WMG, University of Warwick, Coventry, UK

<sup>17</sup>University of Birmingham and Southeast University Joint Research Centre for Biomedical Engineering, Suzhou, China

## Correspondence

Andrew C. Peet, FRCPCH, MBBS, PhD,  
 Institute of Child Health, Whitall Steet,  
 Birmingham, B4 6NH, UK.  
 Email: [a.peet@bham.ac.uk](mailto:a.peet@bham.ac.uk)

MRS can provide high accuracy in the diagnosis of childhood brain tumours when combined with machine learning. A feature selection method such as principal component analysis is commonly used to reduce the dimensionality of metabolite profiles prior to classification. However, an alternative approach of identifying the optimal set of metabolites has not been fully evaluated, possibly due to the challenges of defining this for a multi-class problem. This study aims to investigate metabolite selection from in vivo MRS for childhood brain tumour classification. Multi-site 1.5 T and 3 T

**Abbreviations:** <sup>1</sup>H-MRS, proton MRS; AUC, area under the curve; CRLB., Cramér-Rao lower bound; FWHM, full width at half maximum; LDA, linear discriminant analysis; PCA., principal component analysis; ROC, receiver operating characteristic; SMOTE, synthetic minority oversampling technique; SNR, signal-to-noise ratio.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2022 The Authors. *NMR in Biomedicine* published by John Wiley & Sons Ltd.

### Funding information

Action Medical Research; Cancer Research UK and EPSRC Cancer Imaging Programme; Cancer Research UK and NIHR Experimental Cancer Medicine Centre Paediatric Network; CHILDREN with CANCER UK; Children's Cancer and Leukaemia Group; Children's Research Fund.; Birmingham Women's and Children's Hospital Charities.; The Children's Cancer and Leukaemia Group -- Little Princess Trust; Health Data Research UK; Help Harry Help Others Cancer Charity; NIHR Research Professorship; The Brain Tumour Charity

cohorts of patients with a brain tumour and histological diagnosis of ependymoma, medulloblastoma and pilocytic astrocytoma were retrospectively evaluated. Dimensionality reduction was undertaken by selecting metabolite concentrations through multi-class receiver operating characteristics and compared with principal component analysis. Classification accuracy was determined through leave-one-out and *k*-fold cross-validation. Metabolites identified as crucial in tumour classification include *myo*-inositol ( $P < 0.05$ ,  $AUC = 0.81 \pm 0.01$ ), total lipids and macromolecules at 0.9 ppm ( $P < 0.05$ ,  $AUC = 0.78 \pm 0.01$ ) and total creatine ( $P < 0.05$ ,  $AUC = 0.77 \pm 0.01$ ) for the 1.5 T cohort, and glycine ( $P < 0.05$ ,  $AUC = 0.79 \pm 0.01$ ), total *N*-acetylaspartate ( $P < 0.05$ ,  $AUC = 0.79 \pm 0.01$ ) and total choline ( $P < 0.05$ ,  $AUC = 0.75 \pm 0.01$ ) for the 3 T cohort. Compared with the principal components, the selected metabolites were able to provide significantly improved discrimination between the tumours through most classifiers ( $P < 0.05$ ). The highest balanced classification accuracy determined through leave-one-out cross-validation was 85% for 1.5 T  $^1\text{H}$ -MRS through support vector machine and 75% for 3 T  $^1\text{H}$ -MRS through linear discriminant analysis after oversampling the minority. The study suggests that a group of crucial metabolites helps to achieve better discrimination between childhood brain tumours.

### KEYWORDS

childhood brain tumour, machine learning, metabolite concentration, MRS, multi-class classification, receiver operating characteristics

## 1 | INTRODUCTION

Primary tumours of the central nervous system are the most common cause of cancer death in childhood.<sup>1</sup> Histology and molecular analysis of tumour specimens obtained at operation provide the definitive diagnosis for most children's brain tumours. However, providing an accurate non-invasive diagnosis prior to surgery has many advantages for optimal patient management,<sup>2</sup> including better informed surgery, earlier planning of adjuvant treatment and improved discussions with the child and family. Although clinically used to propose radiological diagnosis, structural features of tumours provided by conventional imaging have a limited accuracy.<sup>3</sup> Metabolites are direct signatures of biochemical activity and their detection *in vivo* can improve our understanding of brain tumours *in situ*.<sup>4</sup> Many studies have suggested specified metabolites as biomarkers of specific processes and types of childhood brain tumour. For instance, total choline, which consists of glycerophosphocholine, phosphocholine and free choline, is known as a general cancer marker and is associated with tumour aggressiveness and progression.<sup>5</sup> Taurine, the naturally occurring  $\beta$ -aminoacid related to neurodevelopment in infancy, is associated with embryonal tumours.<sup>6,7</sup> Lipids and glycine have been identified as valuable prognostic markers of childhood brain tumours.<sup>8-11</sup>

Proton MRS ( $^1\text{H}$ -MRS) is a non-invasive tool to investigate the *in vivo* metabolite profile of tissues in tumours.<sup>12</sup> It has been shown to aid diagnosis and clinical management of childhood brain tumours.<sup>13-15</sup> Machine learning provides a computational method to classify brain tumours by using metabolite profiles and can offer high diagnostic accuracy.<sup>16</sup> Prior to use in a machine learning classifier, a feature selection method is commonly applied to the metabolite profile, to avoid over-fitting by reducing dimensionality. One method that has been commonly used is principal component analysis (PCA), selecting linear combinations of metabolites ranked by their contribution to the overall variability of the data.<sup>17</sup> However, high levels of variability across the whole dataset may not correspond to optimum discrimination between classes, and higher performing methods of feature selection are being sought.<sup>18</sup>

Feature selection methods including PCA commonly have a complex relationship to the original data. This lack of transparency could be a barrier to clinical adoption, where tumour discrimination based on specific metabolites may be preferred by clinicians and can be supported by association with histological features.<sup>19</sup> A feature selection method based on selection of individual metabolites by their ability to discriminate between classes would potentially provide an intuitive approach to moving from diagnosis using single metabolites to machine learning. Whilst diagnostic performance is well defined for classification between two classes by the area under a receiver operating characteristic (ROC) curve,

this is not the case for a multi-class system. Brain tumour classification usually needs to consider more than two potential diagnoses and so multi-class ROC needs to be investigated in this setting. Here we perform a thorough investigation of the accuracy of individual metabolites as discriminators between three major children's brain tumour types comparing different approaches to ROC for a multi-class problem. We then use these approaches to select the best performing metabolites for use in a machine learning classifier. We also compare this approach with feature selection using PCA, which is chosen as a benchmark, although our aim is not to determine whether metabolite selection is the optimal feature selection method when compared with all methods currently available.

The aim of this study is therefore to investigate the use of multi-class ROC for optimizing metabolite selection in the classification of childhood brain tumours by 1.5 T and 3 T short echo-time  $^1\text{H}$ -MRS and compare the classification accuracy against PCA.

## 2 | METHODS

### 2.1 | Data acquisition

Patients with a suspected brain tumour were recruited from four sites in England, including Birmingham Children's Hospital, Alder Hey Children's Hospital Liverpool, Nottingham University Hospitals and Royal Victoria Infirmary Newcastle upon Tyne. Patients presenting with a brain tumour underwent routine MRI examination before the surgical resection of their tumours, and the diagnoses were made by histology and review at the local tumour boards. Patient data were collected from October 2004 to December 2019; those with a diagnosis of pilocytic astrocytoma, medulloblastoma or ependymoma were included in this analysis. The study was approved by the local research ethics committee (ethics number: 04/MRE04/41).

Structural imaging and  $^1\text{H}$ -MRS acquisition were performed on multiple scanners, including Siemens Symphony MAGNETOM 1.5 T and MAGNETOM Verio 3 T (Siemens Healthcare, Erlangen, Germany), GE Signa LX 1.5 T (General Electric Medical Systems, Milwaukee, WI, United States), and Philips Ingenia 1.5 T, Intera 1.5 T, Achieva 3 T, and X-series 3 T (Philips Medical Systems, Eindhoven, The Netherlands), with the appropriate software (Siemens, Symphony Magnetom Syngo MR 2004a, MAGNETOM Verio Syngo MR B17; GE, Signa Excite Hd/x 15.0; Phillips, Ingenia, Intera and Achieva R3.2-5.1). Structural MRI included  $T_1$ -weighted,  $T_2$ -weighted and  $T_1$ -weighted post-contrast sequences as well as diffusion-weighted imaging.  $^1\text{H}$ -MRS with water reference acquisition was performed after conventional imaging that included gadolinium administration by using the point-resolved spectroscopy sequence (field strength 1.5 T or 3 T, head coils or head and neck coils 8-32 channels, sampling frequency 2000-2500 Hz, chemical shift displacement less than 4% per ppm, echo time 30-46 ms, number of complex points 512 or 2048, pulse repetition time 1500-2000 ms, 128 averages collected from a  $20 \times 20 \times 20$ - $80 \times 80 \times 80$  mm<sup>3</sup> volume of interest). Water suppression was performed by using chemical shift selective saturation pulses, and no out of volume suppression was used. Volumes of interest were manually placed to be completely within the tumours according to structural images, with contrast enhancement and low apparent diffusion coefficient being used as guides where tumours exhibited some heterogeneity.<sup>4</sup>

### 2.2 | $^1\text{H}$ -MRS quantification and quality control

All  $^1\text{H}$ -MRS raw data were quantified using TARQUIN (Version 4.3.11),<sup>20</sup> which includes phasing, chemical shift calibration and metabolite amplitude estimation. The basis set used for  $^1\text{H}$ -MRS quantification is a  $^1\text{H}$  brain full basis set that includes lipid basis signals and macromolecule signals. The concentrations obtained were normalized based on the sum of all metabolite, lipid and macromolecule concentrations for each case. Signal-to-noise ratios (SNRs) were calculated with TARQUIN and used to evaluate the level of noise presenting in  $^1\text{H}$ -MRS. Overall SNR was defined as the ratio between the maximum amplitude of the spectrum minus the baseline, and twice the root-mean-square of the residual between 0.2 and 4 ppm. Full width at half maximum (FWHM) was also taken as the value calculated by TARQUIN, and is the width of the unsuppressed water peak at half its full height.

Quality control was then applied in the following manner. Patient data acquired were initially screened according to the following exclusion criteria: (1) histological diagnosis was missing; (2) water suppressed signals, water reference signals or structural MR images indicating the  $^1\text{H}$ -MRS voxel location were missing; (3) the tumour did not occupy all the  $^1\text{H}$ -MRS voxel as determined by visual inspection of the voxel location images produced on the scanner aided by reference to the available image set; (4) spectra showed very poor FWHM ( $>0.15$  ppm, equivalent to 9.6 Hz in 1.5 T and 19.2 Hz in 3 T); (5) spectra showed very poor SNR ( $<4$ ). For the cases that met these metric-based quality measures, the TARQUIN-processed frequency domain  $^1\text{H}$ -MRS of all cases was assessed visually by experienced spectroscopists for general quality features, namely phasing, fitting, baseline variation and the presence of artefacts.

## 2.3 | Metabolite evaluation

Cramér-Rao lower bounds (CRLBs) were used to evaluate the uncertainty of metabolite quantification. Metabolites whose CRLB as a percentage was greater than 50%, for all cases in both 1.5 T and 3 T cohorts, were excluded. This eliminated alanine, aspartate,  $\gamma$ -aminobutyric acid and phosphatidylethanol from the analysis in both 1.5 T and 3 T cohorts. Multi-class diagnostic performance of individual metabolites, for all tumours, was evaluated using multi-class ROC.<sup>21</sup> The area under the curve (AUC) for a multi-class ROC is defined by

$$AUC_{\text{multi-class}} = \frac{2}{c(c-1)} \sum AUC_{\text{binary}} = \frac{2}{c(c-1)} \sum \frac{1}{2} (\hat{A}(ij) + \hat{A}(ji)),$$

where  $AUC_{\text{binary}}$  denotes the AUC of the binary ROC,  $c$  denotes the number of classes and  $\hat{A}(ij)$  denotes the probability that a randomly selected element of class  $j$  will have a lower estimated probability of belonging to class  $i$  than a randomly selected element of class  $i$ .

Multi-class diagnostic performance was compared with the binary diagnostic performance given by binary and pairwise ROC. For binary problems, the multi-class ROC was converted into a two-class problem by considering class A and class non-A. In pairwise ROC, two of the three tumour types were selected, and the other type was ignored. The standard deviation of then AUC was measured through leave-one-out cross-validation. Metabolites were selected based on the diagnostic ability for the childhood brain tumours in this study derived using multi-class ROC.

## 2.4 | Tumour classification

The <sup>1</sup>H-MRS analysis (Figure 1) and tumour classification were conducted for the 1.5 T and 3 T cohorts separately. By taking account of the known challenges of measuring certain metabolites with high levels of spectral overlap,<sup>22</sup> a group of 14 metabolites was selected to be ranked for diagnostic ability, including the citrate, glutathione, glycine, lactate, *myo*-inositol, *scyllo*-inositol, taurine, total *N*-acetylaspartate (*N*-acetylaspartate and *N*-acetylaspartylglutamate), total choline (glycerophosphocholine and phosphocholine), total creatine (creatine and phosphocreatine), combined glutamate and glutamine, and total lipids and macromolecules at 0.9 ppm, 1.3 ppm and 2.0 ppm.

Training sets and test sets were sampled from the whole set, with stratification according to tumour types. Features were extracted only from the training set and for building classifiers. Two methods of feature extraction were individually performed and compared, including PCA and multi-class ROC. The number of features was determined as the sample size of the minority group minus one. For the method based on PCA, principal components were derived by performing PCA on the matrix of all screened metabolites and ranked based on the explained cumulative variance. For the method of multi-class ROC, metabolites were ranked based on the AUC derived through multi-class ROC. Highly ranked metabolites or principal components were used as features for tumour classification.

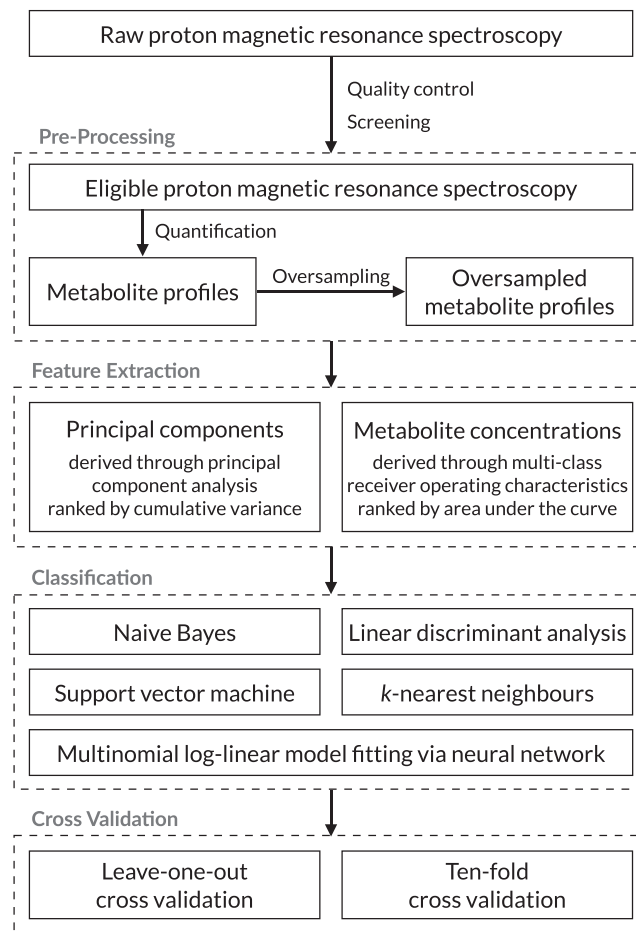
Ependymomas were oversampled by 100% through the adaptive synthetic minority oversampling technique (SMOTE), in order to correct the skewness and class imbalance of the cohort.<sup>23</sup> Linear and non-linear classifiers were applied to evaluate the classification performance, including linear discriminant analysis (LDA), *k*-nearest neighbours, naïve Bayes, multinomial log-linear model fitting via a neural network, and support vector machine with a linear kernel. Discriminant functions derived through LDA and re-substitution were used to show poorly classified cases with low classification probability. Leave-one-out and *k*-fold cross-validation was used to determine classification accuracy, where *k* was determined based on the size of the minority class. The *k*-fold cross-validation is usually more accurate,<sup>24</sup> but it may lead to poor training sets being selected when the cohort is small and particularly when the minority class is very small. Both cross-validation methods were therefore performed to achieve some level of comparability, whilst not selecting a more appropriate method for either cohort. Overall ( $\alpha_{\text{overall}}$ ) and balanced ( $\alpha_{\text{balanced}}$ ) classification accuracies were used to evaluate the classification performance, defined based on the accuracy for each case ( $\alpha_i$ ) as

$$\alpha_{\text{overall}} = \frac{1}{N_{\text{cases}}} \sum_{i=1}^{N_{\text{cases}}} \alpha_i, \quad (1)$$

$$\alpha_{\text{balanced}} = \frac{1}{3} \left( \frac{1}{N_{\text{EP}}} \sum_{i=1}^{N_{\text{EP}}} \alpha_i + \frac{1}{N_{\text{MB}}} \sum_{i=1}^{N_{\text{MB}}} \alpha_i + \frac{1}{N_{\text{PA}}} \sum_{i=1}^{N_{\text{PA}}} \alpha_i \right). \quad (2)$$

## 2.5 | Statistical analysis

A Kruskal-Wallis *H* test was performed to assess the different quality metrics, the means of metabolite concentrations across the three tumour types, and the classification accuracy. Statistically significant differences between tumour types or processing methods were determined when



**FIGURE 1** Flowchart of the method for performing  $^1\text{H}$ -MRS analysis

$P < 0.05$ ,  $P < 0.01$ ,  $P < 0.001$  and  $P < 0.0001$ . All algorithms of statistical analysis, feature extraction and machine learning were implemented using R (Version 3.6.2, R Foundation, Vienna, Austria).

### 3 | RESULTS

#### 3.1 | Demographics

Diagnostic  $^1\text{H}$ -MRS was performed on 116 patients on 1.5 T and 73 patients on 3 T MR scanners. Eighty-three (66%) 1.5 T cases and 42 (34%) 3 T cases were enrolled after screening of data availability and quality control assessment (Table 1, Figure 2). Tumours were generally located in the posterior fossa, but four ependymomas were located supratentorially, including three for 1.5 T and one for 3 T. Ages of patients ranged from 1.8 months to 18 years of age, across the three groups. Forty-seven (57%) of the 1.5 T cases and 18 (43%) of the 3 T cases were male. Histological subtypes of all tumours were grouped together within each tumour type (Table 2).

#### 3.2 | Quality assessment

Accepted cases in the final cohort showed overall SNR and FWHM as  $19 \pm 13$  and  $5 \pm 1$  Hz for the 1.5 T cohort (Figure 2A), and  $16 \pm 11$  and  $7 \pm 2$  Hz for the 3 T cohort (Figure 2B and Table 3). Medulloblastoma cases showed a generally better signal quality than did the other two groups ( $P < 0.05$ ). Median CRLB of the metabolites ranged from 18% to 299% for the 1.5 T cohort, and 21% to 247% for the 3 T cohort.

**TABLE 1** The multi-site cohorts of  $^1\text{H}$ -MRS for childhood brain tumours

1.5 T cohort	Ependymomas (16%)	Medulloblastomas (37%)	Pilocytic astrocytomas (47%)	Total
Initial sample	23	39	54	116
Missing histology	0	1	0	1
Incomplete imaging data	2	4	3	9
Partial volume effects	1	1	3	5
Poor line width	3	0	1	4
Poor SNR	2	0	2	4
Artefacts	2	2	6	10
Final sample	13*	31	39	83
Birmingham (96%)	10	31	39	80
Liverpool (1%)	1	0	0	1
Nottingham (3%)	2	0	0	2
3 T cohort	Ependymomas (10%)	Medulloblastomas (40%)	Pilocytic astrocytomas (50%)	Total
Initial sample	7	25	41	73
Missing histology	0	2	1	3
Incomplete imaging data	2	4	6	12
Partial volume effects	1	0	4	5
Poor line width	0	0	0	0
Poor SNR	0	0	1	1
Artefacts	0	2	8	10
Final sample	4**	17	21	42
Birmingham (53%)	3	8	11	22
Liverpool (38%)	1	8	7	16
Newcastle (7%)	0	1	2	3
Nottingham (2%)	0	0	1	1

\*Three of the 13 ependymomas were located supratentorially.

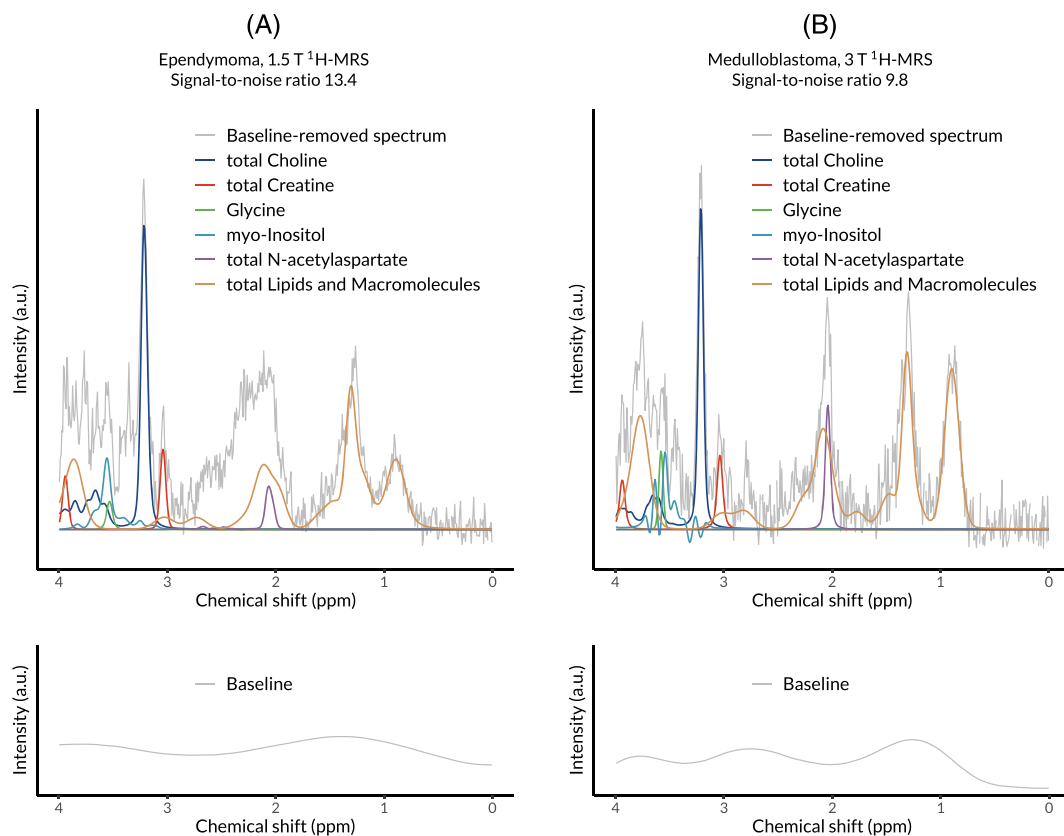
\*\*One of the four ependymomas was located supratentorially.

### 3.3 | Univariate analysis of metabolites

Average  $^1\text{H}$ -MRS spectra showed visual differences among the three tumour types (Figure 3). Acquired  $^1\text{H}$ -MRS showed key metabolites, with the mean values differing significantly between the three brain tumour types (Tables 4) and having diverse diagnostic ability (Figure 4). The AUC values of the top four metabolites for the 1.5 T and 3 T cohorts are shown in Figure 5, where multi-class ROC is compared with the various binary and pairwise ROCs. As expected, some of these metabolites have poor discriminatory ability for specific tumours, reflected in low binary or pairwise AUC, but the multi-class AUC provides a good reflection of the overall performance for each metabolite. Several metabolites showed good diagnostic ability (AUC > 0.7) at both of the field strengths. In the 1.5 T cohort (Figure 4A), *myo*-inositol (Figure 5b) was ranked as the most discriminatory metabolite, followed by total lipids and macromolecules at 0.9 ppm (Figure 5B), total creatine (Figure 5C) and total *N*-acetylaspartate (Figure 5D). In the 3 T cohort (Figure 4B), total *N*-acetylaspartate (Figure 5E) was ranked as the most discriminatory metabolite, followed by glycine (Figure 5F), total choline (Figure 5G) and taurine (Figure 5H). Compared with the results of 1.5 T  $^1\text{H}$ -MRS, several metabolites showed clearly improved diagnostic ability in the 3 T  $^1\text{H}$ -MRS, including glycine, total *N*-acetylaspartate, total choline, taurine, and total lipids and macromolecules at 2.0 ppm. However, some other metabolites showed decreased diagnostic ability, including combined glutamate and glutamine, lactate, *myo*-inositol, total lipids and macromolecules at 0.9 ppm, and total creatine (Figure 4A and 4B).

### 3.4 | Principal component analysis

The results of PCA showed that similar numbers of principal components accounted for the same proportions of the total variance in metabolite profiles in the 1.5 T and 3 T  $^1\text{H}$ -MRS (Figure 4). Four principal components were able to explain around half of the total variance, which is also the



**FIGURE 2** Images showing quantification examples with the corresponding baseline under the same scale for the acquired 1.5 T and 3 T  $^1\text{H}$ -MRS for childhood brain tumours, an ependymoma case from the 1.5 T cohort (A) and a medulloblastoma cases from the 3 T cohort (B)

maximum number of principal components allowed to be used in 3 T  $^1\text{H}$ -MRS before oversampling the minority class (Figure 4C and 4D). Eleven principal components were able to account for around 95% of the variance and were used in feature extraction for oversampled 1.5 T  $^1\text{H}$ -MRS (Figure 4C). Meanwhile, seven principal components were available in oversampled 3 T  $^1\text{H}$ -MRS, and they were able to explain around 70% of the total variance (Figure 4D).

### 3.5 | Classification performance

Classification performance was evaluated using classification accuracy, with a further evaluation of misclassified cases. Poorly classified cases were compared between the two feature extraction methods (Figures 6). Machine learning showed limited ability in classifying tumours that were not representative of their diagnostic types. Oversampling for the minority class showed assistance for classifying the ependymomas.

Significantly improved classification accuracy was obtained by using selected metabolites through multi-class ROC, compared with PCA-based feature selection ( $P < 0.01$ , Figure 7A and 7B). The combination of multi-class ROC and oversampling showed further improved classification accuracy ( $P < 0.01$ , Figure 7C and 7D). The improvement was seen for both overall and balanced classification accuracy ( $P < 0.01$ ) and remained significant through  $k$ -fold cross-validation ( $P < 0.01$ , Figure 8). The improvement of classification performance was consistent between different classifiers in 1.5 T (Supplementary Figures 1 and 3) and 3 T (Supplementary Figures 2 and 4)  $^1\text{H}$ -MRS, which is validated through leave-one-out cross-validation (Supplementary Figures 1 and 2) and  $k$ -fold cross-validation (Supplementary Figures 3 and 4). Optimal classification accuracy was achieved through a support vector machine and oversampling for the 1.5 T  $^1\text{H}$ -MRS as overall accuracy of 88% and balanced accuracy of 85%. At the same time, LDA and oversampling provided the best classification accuracy for the 3 T  $^1\text{H}$ -MRS as overall accuracy of 84% and balanced accuracy of 75%.

## 4 | DISCUSSION

In this study, the role of metabolite selection for optimizing childhood brain tumour classification from 1.5 T and 3 T short echo-time single-voxel  $^1\text{H}$ -MRS has been investigated. A method of metabolite selection in childhood brain tumour classification through multi-class ROC was presented.

**TABLE 2** Demographic and clinical variables of patients

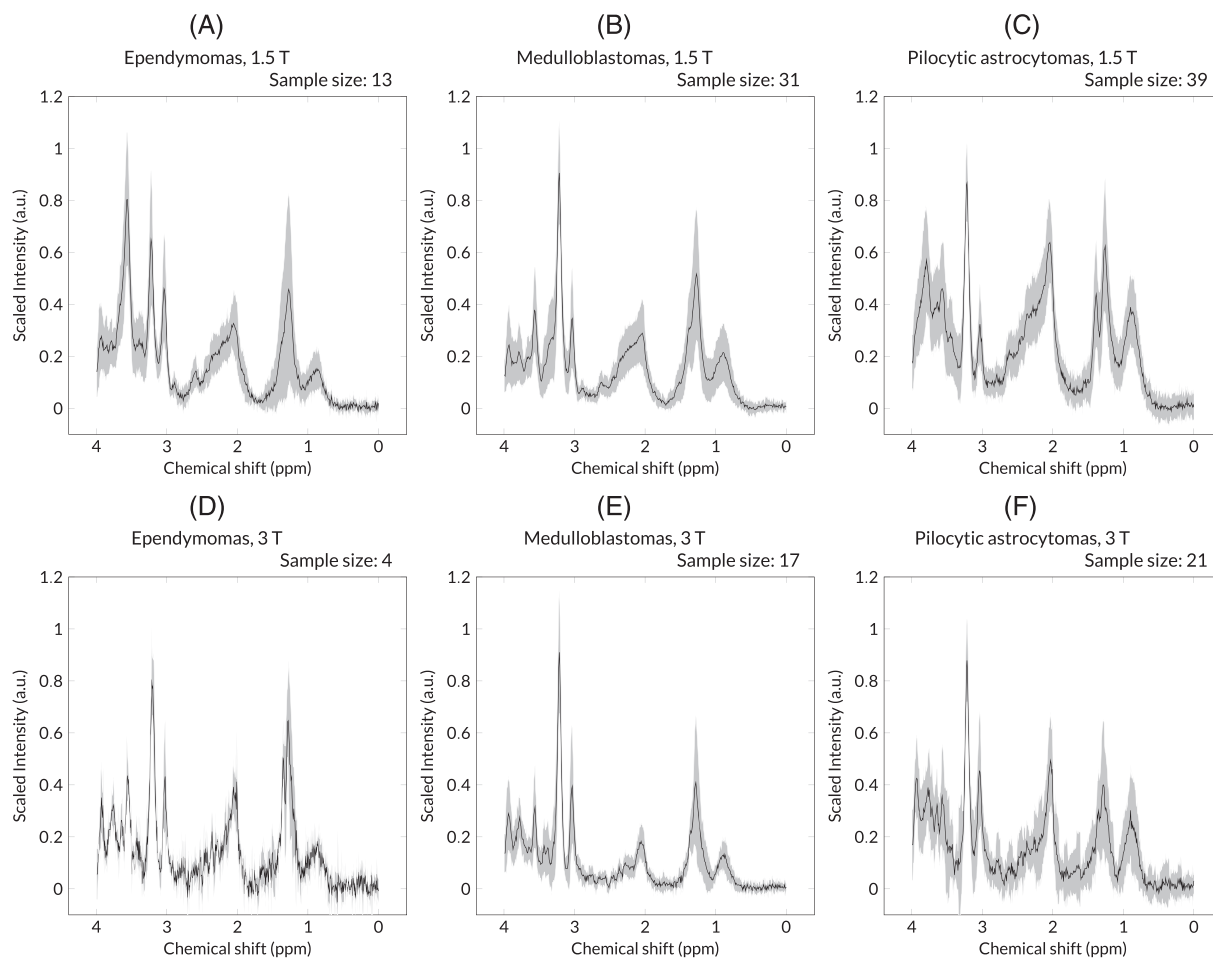
Sample size	1.5 T cohort 83	3 T cohort 42
Site of primary tumours		
Ependymomas	13	4
Medulloblastomas	31	17
Pilocytic astrocytomas	39	21
Age in years		
At diagnosis		
Ependymomas	4.6 ± 5.3	9.2 ± 5.3
Medulloblastomas	6.9 ± 3.5	6.5 ± 4.4
Pilocytic astrocytomas	8.0 ± 3.6	6.3 ± 5.1
At image acquisition		
Ependymomas	4.6 ± 5.3	9.3 ± 5.4
Medulloblastomas	6.9 ± 3.5	6.5 ± 4.4
Pilocytic astrocytomas	8.0 ± 3.6	6.3 ± 5.0
Gender		
Male		
Ependymomas	5	0
Medulloblastomas	22	10
Pilocytic astrocytomas	20	8
Female		
Ependymomas	6	4
Medulloblastomas	11	7
Pilocytic astrocytomas	19	13
Histological subtypes		
Ependymomas		
Grade II	7	2
Grade III, anaplastic	6	2
Medulloblastomas		
Grade IV, classic	26	17
Grade IV, desmoplastic-nodular	5	0
Pilocytic astrocytomas		
Grade I	39	21

**TABLE 3** Quality metrics of the accepted <sup>1</sup>H-MRS

		Overall SNR	FWHM (Hz)
1.5 T cohort	Ependymomas	21 ± 7	5 ± 1
	Medulloblastomas	30 ± 16	4 ± 1
	Pilocytic astrocytomas	11 ± 4	5 ± 1
3 T cohort	Ependymomas	10 ± 4	7 ± 2
	Medulloblastomas	25 ± 12	6 ± 2
	Pilocytic astrocytomas	10 ± 3	7 ± 2

Across the commonly used machine learning methods, this method is able to achieve an improved classification accuracy for the three main tumour types when compared with the conventional PCA for feature selection. A combination of metabolite selection through multi-class ROC and oversampling for the minority through SMOTE is able to achieve the optimal classification accuracy, providing an accurate, efficient and transparent method for the use of metabolites in children's brain tumour diagnosis.





**FIGURE 3** Images showing the average of scaled and baseline-removed spectrum for ependymomas (A, D), medulloblastomas (B, E) and pilocytic astrocytomas (C, F) in the 1.5 T (A-C) and 3 T (D-F)  $^1\text{H}$ -MRS cohorts, which are presented as means with standard deviations

With the focus of the feature selection being the metabolite concentrations, careful consideration of the most appropriate metabolite set is important. All metabolites in the basis set were initially quantified separately in TARQUIN as part of its standard application. Some metabolites were then combined. Creatine and phosphocreatine have  $^1\text{H}$ -MRS signals that almost completely overlap at 1.5 T and 3 T, and we combine them. Phosphocholine, glycerophosphocholine and free choline all have spectra that are dominated by a singlet around 3 ppm, and whilst potentially they could be separated by their multiplets between 3 ppm and 4 ppm, these are of low intensity and in tumours this region of the spectrum has signals from many metabolites, so we decided to combine them. Glycine and *myo*-inositol differ from the creatine and choline containing metabolite sets in that their molecular structures and spectra are substantially different from each other, and whilst there is significant overlap in the spectra of the two metabolites, particularly at 1.5 T, this gives the opportunity to quantify them separately. Previous publications on children's brain tumour  $^1\text{H}$ -MRS have shown that glycine and *myo*-inositol can be quantified separately, albeit with limited accuracy, and provide useful information using LCModel<sup>8</sup> and TARQUIN<sup>11</sup> with short echo-time point-resolved spectroscopy at 1.5 T. We would expect glycine and *myo*-inositol to be quantified more accurately with increasing field strength, since the complex multiplet structure of *myo*-inositol becomes more evident, reducing the overlap in the spectra.<sup>25</sup>

It is important to have some comparison of metabolite selection against a commonly used method, and PCA was chosen, as it has been used successfully on similar datasets.<sup>17</sup> It is an unsupervised learning method with categories of cases not considered, and so is valid for use in highly imbalanced data, a situation commonly encountered in children's brain tumour classification, since some tumour types are more common than others. However, the features are ranked by their variability across the whole cohort and highly variable features may not be the best at discriminating between classes. In contrast, selecting metabolites by their ability to discriminate between the various classes and using these as the features for machine learning should improve feature selection, since it is a supervised method. ROC is well defined between pairs of classes, but its generalization to the multi-class problem is less commonly used in biomedical applications. The multi-class problem can be reduced to a series of binary problems either by selecting pairs of classes or in a one versus the rest approach. However, this produces multiple metrics for each metabolite. This is well illustrated in Figure 5, where it is seen that the binary and pairwise AUC values show considerable variability across the methods.

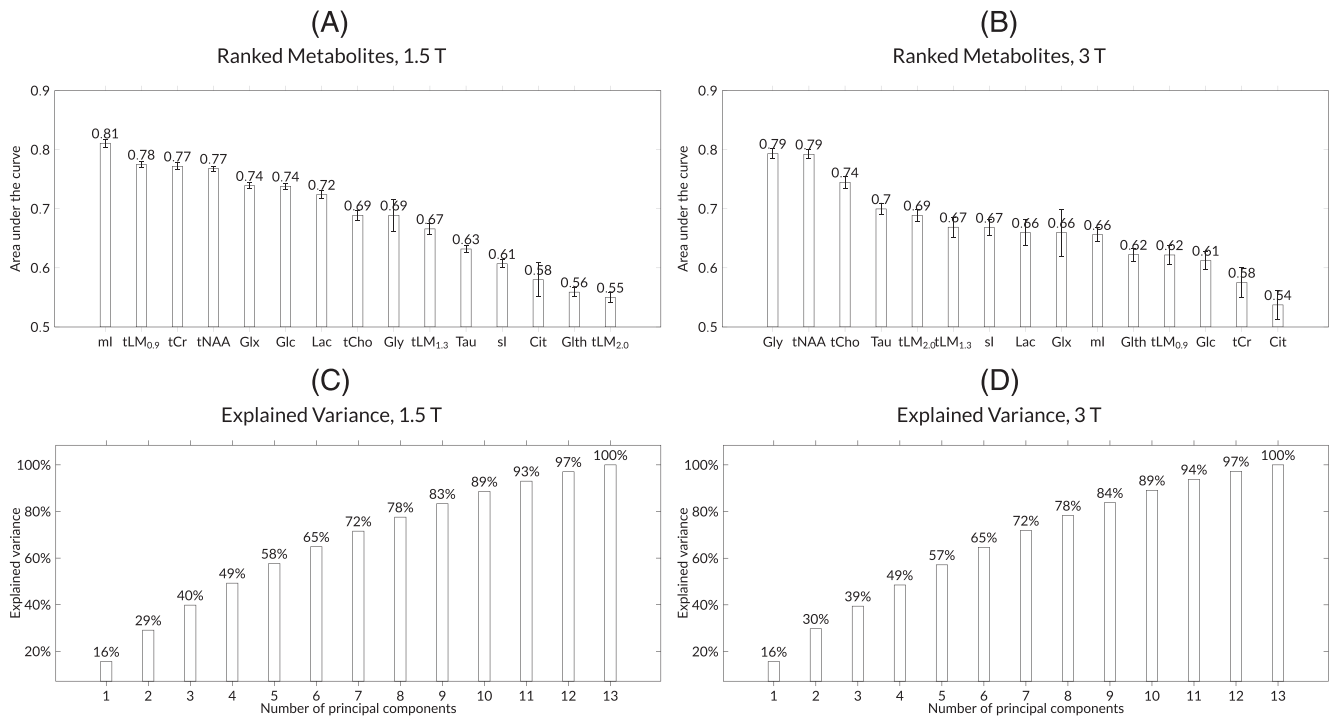
**TABLE 4** Estimated metabolite concentrations of childhood brain tumours

Metabolites	EP Mean ± SD (mM)	MB Mean ± SD (mM)	PA Mean ± SD (mM)	All Mean ± SD (mM)	P EP MB PA	
<b>1.5 T cohort</b>	<b>N = 13</b>	<b>N = 31</b>	<b>N = 39</b>	<b>N = 83</b>		
Citrate	0.7 ± 0.3	0.6 ± 0.3	0.4 ± 0.3	0.5 ± 0.3	<0.001	**
Total choline*	1.9 ± 1.0	3.3 ± 1.6	1.1 ± 0.4	2.0 ± 1.5	<0.001	**
Total creatine**	3.9 ± 1.8	2.9 ± 1.7	0.9 ± 1.2	2.1 ± 1.9	<0.001	**
Glucose	1.2 ± 0.9	0.6 ± 0.9	1.0 ± 0.8	0.9 ± 0.9	0.005	**
Combined glutamate and glutamine	6.9 ± 2.4	6.4 ± 2.5	5.7 ± 2.2	6.1 ± 2.4	0.150	
Glutathione	0.4 ± 0.5	0.7 ± 0.5	0.3 ± 0.3	0.4 ± 0.4	0.010	**
Glycine	1.9 ± 3.5	3.1 ± 2.5	0.2 ± 0.3	1.5 ± 2.5	<0.001	**
Lactate	1.7 ± 1.7	2.6 ± 1.6	1.9 ± 1.1	2.1 ± 1.4	0.048	*
Total lipids and macromolecules at 0.9 ppm	4.3 ± 2.6	7.4 ± 4.1	4.0 ± 1.5	5.3 ± 3.3	<0.001	**
Total lipids and macromolecules at 1.3 ppm	14.9 ± 16.8	20.4 ± 15.2	7.3 ± 4.7	13.4 ± 13.1	<0.001	**
Total lipids and macromolecules at 2.0 ppm	7.9 ± 2.1	9.6 ± 3.1	5.2 ± 1.9	7.2 ± 3.2	<0.001	**
myo-inositol	9.6 ± 5.7	1.7 ± 1.7	2.0 ± 1.8	3.1 ± 3.9	<0.001	**
total N-acetylaspartate***	1.2 ± 0.9	0.9 ± 0.7	1.3 ± 0.8	1.2 ± 0.8	0.067	
scyllo-inositol	0.2 ± 0.3	0.3 ± 0.4	0.1 ± 0.1	0.2 ± 0.3	<0.001	**
Taurine	1.8 ± 1.6	3.1 ± 2.7	0.8 ± 1.1	1.8 ± 2.2	<0.001	**
<b>3 T cohort</b>	<b>N = 4</b>	<b>N = 17</b>	<b>N = 21</b>	<b>N = 42</b>		
Citrate	0.5 ± 0.6	0.7 ± 0.3	0.4 ± 0.3	0.5 ± 0.4	0.080	
Total choline	1.3 ± 0.4	4.5 ± 3.7	1.7 ± 0.9	2.8 ± 2.8	<0.001	**
Total creatine	1.9 ± 1.1	4.4 ± 2.8	3.1 ± 2.2	3.5 ± 2.5	0.027	*
Glucose	0.9 ± 0.3	1.5 ± 1.2	0.6 ± 0.7	1.0 ± 0.1	0.110	
Combined glutamate and glutamine	3.5 ± 0.1	4.8 ± 3.8	5.1 ± 3.6	4.8 ± 3.5	0.590	
Glutathione	0.9 ± 0.3	2.2 ± 2.1	0.9 ± 0.8	1.4 ± 1.5	0.002	**
Glycine	1.2 ± 0.8	3.9 ± 2.3	0.9 ± 1.5	2.2 ± 2.3	<0.001	**
Lactate	1.7 ± 0.8	2.2 ± 1.9	0.9 ± 1.1	1.5 ± 1.5	0.016	*
Total lipids and macromolecules at 0.9 ppm	3.6 ± 1.3	6.4 ± 6.2	5.0 ± 5.7	5.4 ± 5.6	0.050	*
Total lipids and macromolecules at 1.3 ppm	13.3 ± 10.8	23.0 ± 18.3	10.3 ± 10.9	15.7 ± 15.3	<0.001	**
Total lipids and macromolecules at 2.0 ppm	6.2 ± 2.2	8.7 ± 4.1	6.8 ± 3.7	7.5 ± 3.8	0.076	
myo-inositol	2.9 ± 2.4	2.7 ± 2.6	3.0 ± 3.9	2.9 ± 3.2	0.970	
total N-acetylaspartate	1.0 ± 0.5	1.2 ± 1.1	1.9 ± 1.1	1.6 ± 1.1	0.008	**
scyllo-inositol	0.0	0.4 ± 0.7	0.1 ± 0.3	0.2 ± 0.5	0.012	*
Taurine	1.0 ± 0.7	5.7 ± 4.6	1.1 ± 0.3	3.0 ± 0.5	<0.001	**

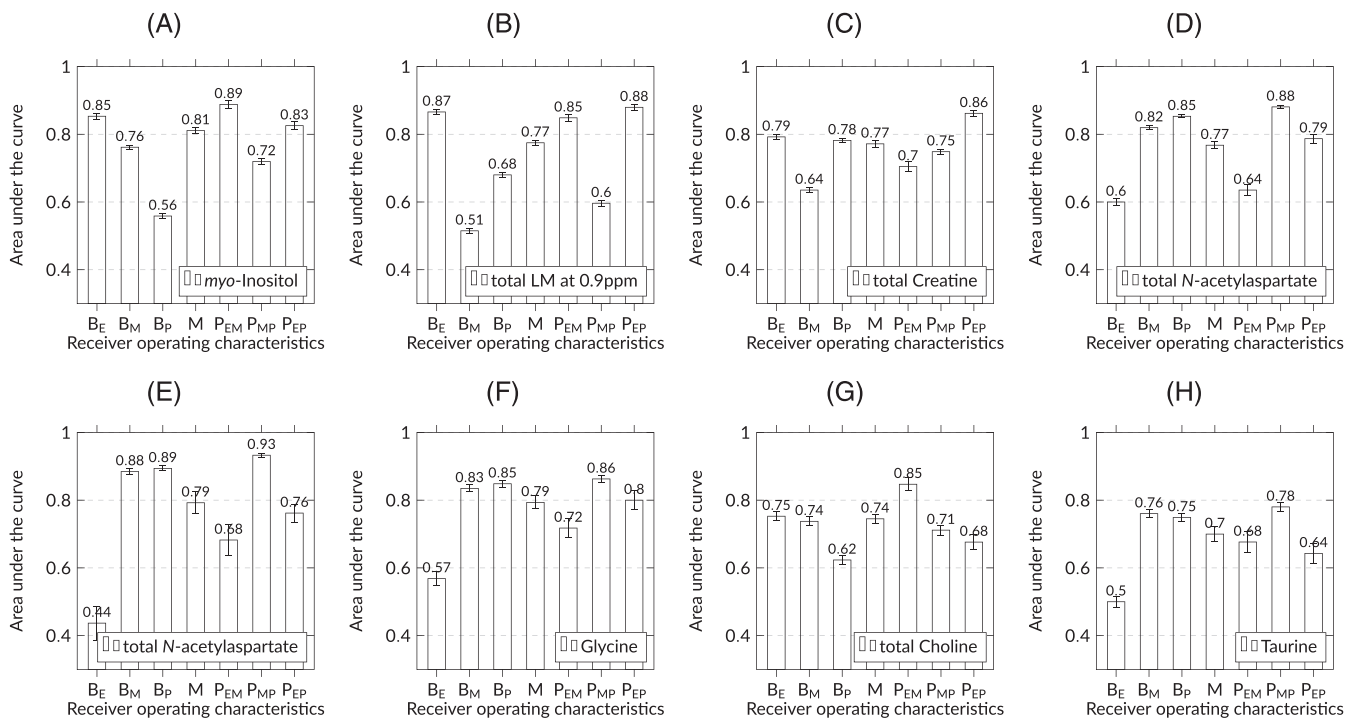
Metabolite concentrations prior to normalization are shown above. *P* values were calculated using non-parametric Kruskal-Wallis *H* test between ependymomas (EP), medulloblastomas (MB) and pilocytic astrocytomas (PA). Significance of differences for metabolite concentrations between the three tumour types is identified as \**P* < 0.05 and \*\**P* < 0.01. For the metabolite combinations, total choline includes glycerophosphocholine and phosphocholine, total creatine includes creatine and phosphocreatine and total *N*-acetylaspartate includes *N*-acetylaspartate and *N*-acetylaspartylglutamate.

Most metabolites are better at diagnosing some tumours than others; for example, *myo*-inositol (Figure 5A) is seen as a good discriminator of ependymomas in the 1.5 T cohort by high binary and pairwise AUC values, but it has a low ability to discriminate between medulloblastomas and pilocytic astrocytomas on the pairwise ROC. Here we use a single combined multi-class ROC parameter, which is contributed equally by all tumour groups and removes this challenge to metabolite selection. One advantage of using the metabolites directly as features is that metabolite concentrations are meaningful in tumour biology and can be compared with recent findings from biopsy studies, which helps to validate their use as features.<sup>26</sup>

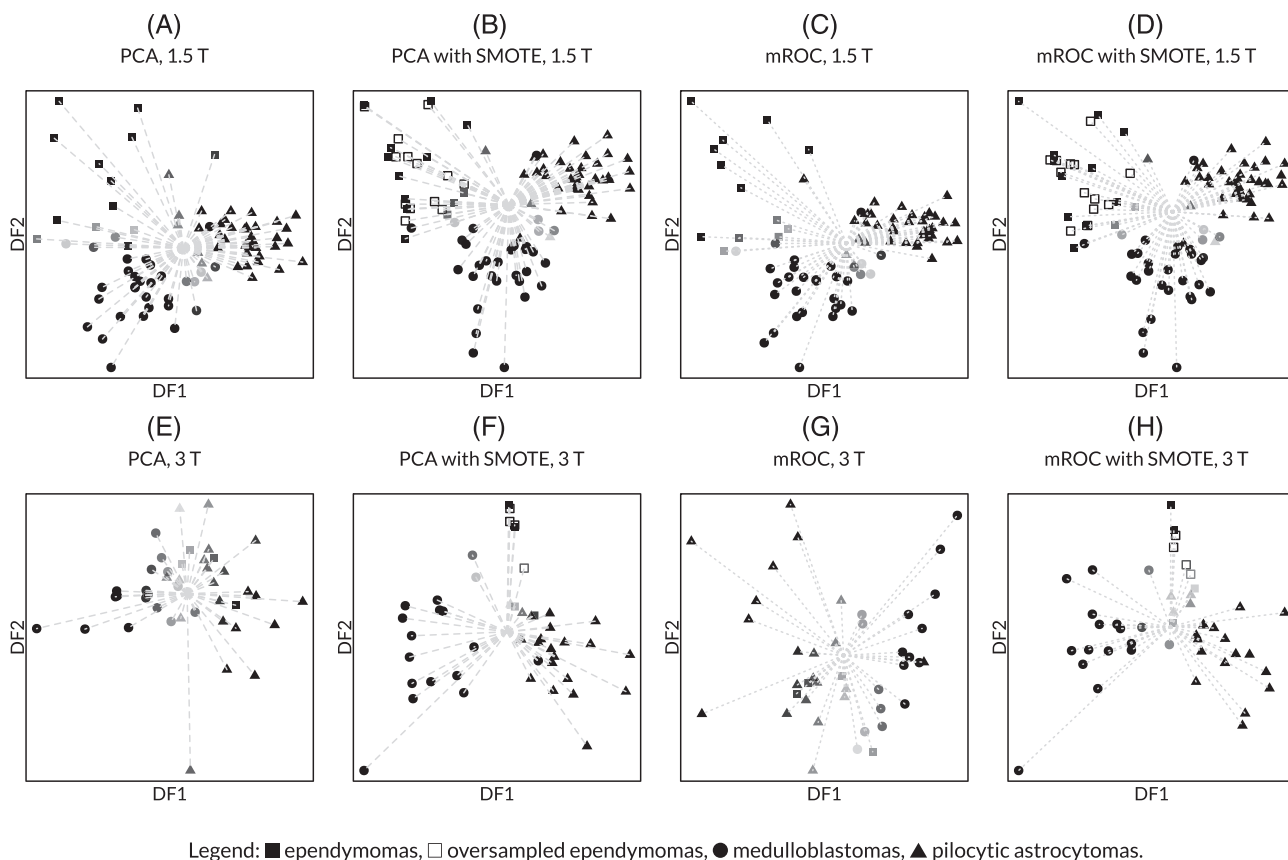
Metabolites have been identified that are characteristic for specific tumours.<sup>17</sup> For instance, previous studies suggested that posterior fossa pilocytic astrocytomas have significantly altered levels of creatine, choline, glutamate, glutamine and *myo*-inositol.<sup>27,28</sup> Glycine was identified as a key metabolite for classifying high-grade and low-grade childhood brain tumours.<sup>8</sup> These findings correspond to the selected metabolites in our



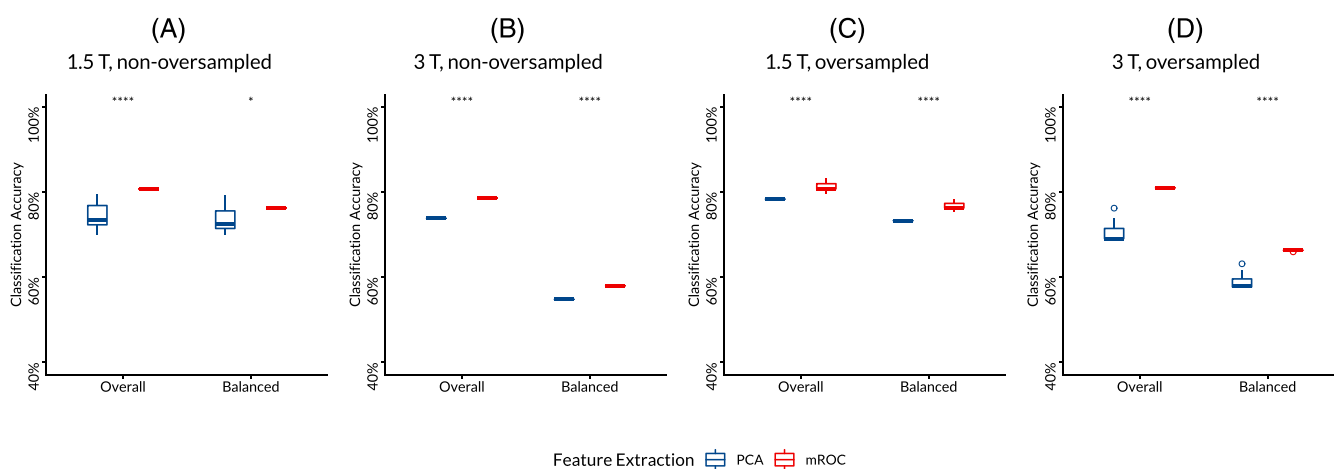
**FIGURE 4** Ranking of metabolites (A, B) and explained variance (C, D) of the 1.5 T (A, C) and 3 T (B, D) <sup>1</sup>H-MRS cohorts. Abbreviations: Cit, citrate; Glc, glucose; GltH, glutathione; Glx, combined glutamate and glutamine; Gly, glycine; Lac, lactate; ml, myo-inositol; tCho, total choline; tCr, total creatine; tNAA, total *N*-acetylaspartate; sl, *scyllo*-inositol; Tau, taurine; tLM<sub>0.9</sub>/tLM<sub>1.3</sub>/tLM<sub>2.0</sub>, total lipids and macromolecules at 0.9/1.3/2.0 ppm



**FIGURE 5** Diagnostic ability of top four metabolites determined from the 1.5 T (A-D) and 3 T (E-H) <sup>1</sup>H-MRS cohorts. The AUC is given as mean and standard deviation, derived by using binary, multi-class and pairwise ROC. Abbreviations: LM, lipids and macromolecules; B<sub>X</sub>, binary tumour X against non-X; M, multi-class; P<sub>XY</sub>, pairwise tumour X against tumour Y; E, ependymomas; M, medulloblastomas; P, pilocytic astrocytomas

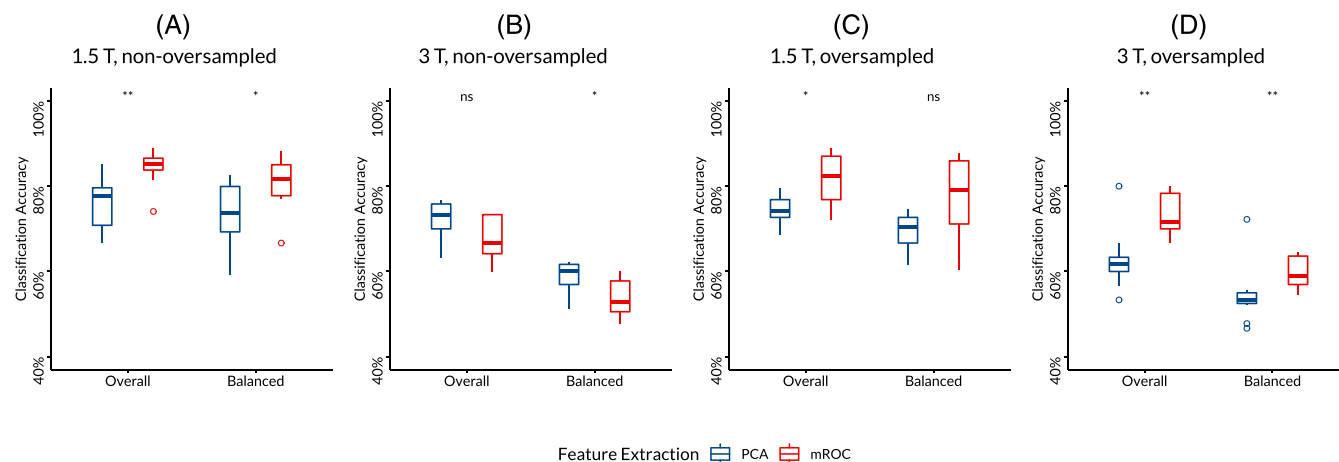


**FIGURE 6** Two-dimensional scatter plots showing the classification performance through re-substitution and  $N_{minority} - 1$  principal components (A, B, E, F) or metabolites (C, D, G, H) for the 1.5 T (A-D) and 3 T (E-H)  $^1\text{H}$ -MRS, also comparing the non-oversampled (A, C, E, G) and oversampled (B, D, F, H) cohorts, where uncertainly classified cases are shown as transparent, with the contrast indicating the probability of classification. Abbreviations: DF, discriminant function; mROC, multi-class receiver operating characteristic



**FIGURE 7** Images showing the classification accuracy determined through multinomial log-linear model fitting via neural networks and leave-one-out cross-validation, for non-oversampled (A, B) and oversampled (C, D) 1.5 T (A, C) and 3 T (B, D)  $^1\text{H}$ -MRS through SMOTE, where both overall and balanced classification accuracy are compared and significance is marked as \* ( $P < 0.05$ ) and \*\*\*\* ( $P < 0.0001$ ). Abbreviations: ns, not significant

results. However, some metabolites were found to be less useful for discriminating between certain tumour types. For instance, total lipids and macromolecules at 0.9 ppm performed well in discriminating between ependymomas and pilocytic astrocytomas but were less useful for identifying medulloblastomas (Figure 5B). This finding demonstrated the benefits of selecting key metabolites for tumour classification.



**FIGURE 8** Images showing the classification accuracy determined through multinomial log-linear model fitting via neural networks and  $k$ -fold cross-validation, for non-oversampled (A, B) and oversampled (C, D) 1.5 T (A, C) and 3 T (B, D)  $^1\text{H}$ -MRS through SMOTE, where both overall and balanced classification accuracy are compared and significance is marked as \* ( $P < 0.05$ ) and \*\* ( $P < 0.01$ )

Comparing the metabolite selection from the 1.5 T and 3 T cohorts, a small group of metabolites showed useful discriminators at both field strengths, including total  $N$ -acetylaspartate, total choline, and total lipids and macromolecules at 1.3 ppm. In this study, glycine was suggested as the most discriminatory biomarker for classification from the 3 T cohort, whereas *myo*-inositol was more important in the 1.5 T cohort. This may be a result of glycine being more accurately discriminated from *myo*-inositol at 3 T than at 1.5 T, as these metabolites have very similar chemical shifts. There are also some other metabolites showing increased diagnostic ability from the 1.5 T to the 3 T cohort, including *scyllo*-inositol and taurine. Again, this could be due to the more accurate metabolite estimation in higher field-strength scanning. There are also some other metabolites performing with a varied diagnostic ability dependent on the field strength, such as combined glutamate and glutamine in the 3 T cohort and citrate in the 1.5 T cohort. Some metabolites are easier to quantify with  $^1\text{H}$ -MRS at 3 T than at 1.5 T due to  $J$ -coupling effects reducing spectral overlap; however, there are some metabolites that have different diagnostic abilities at the two field strengths not readily explained by this. We note that  $T_1$  and  $T_2$  values vary between metabolites and with field strength, which will affect quantification.<sup>29</sup>

The ROC was originally proposed for binary classification problems as applied to radiology research.<sup>30</sup> Its implementation for multi-class classification is still being debated. Our cohort contains imbalanced data, due to the rareness of ependymomas, which makes the problem even more complicated. The multi-class ROC method, employed here, converges with the binary ROC when the problem is converted from multiple to two classes. After evaluating the performance of ROC in binary, pairwise and multi-class evaluation, the results of multi-class ROC performed like a combination of binary and pairwise evaluations, making it ideal for situations where there are multiple classes.

Metabolite selection was used in the current study in combination with number of machine learning classifiers. Whilst the list of classifiers was not exhaustive, major types were included and the purpose was to show the robustness of metabolite selection across a range of classifiers rather than to determine the optimum method. Some insight can be gained from the differences in the results between them. Taking the difference between SVM and LDA as an example, SVM would allow the boundary to be specified more precisely in regions where two classes have neighbours that are very close, while LDA provides a straight boundary, which may work poorly for neighbouring cases. This may well explain why SVM outperforms LDA in the 1.5 T dataset, which has several neighbouring cases from different tumour types. In the 3 T dataset, the small number of ependymomas have disparate metabolite profiles, which overlap other tumour types, leading to poor performance of all machine learning methods for this class.

The limitations of this study include the relatively small cohort sizes, particularly at 3 T, and the challenges in quantifying metabolites from  $^1\text{H}$ -MRS acquired clinically at multiple centres. The cohorts used in the 1.5 T and 3 T analyses are different; indeed, they are mutually exclusive and caution should be exercised in any comparison of results between the two cohorts. The data are also acquired from multiple centres with some variation in protocol, which will lead to variability in the data. Many confounders may be present in addition to the effects of field strength. We have given some detail for both the patients and the methods to aid comparison. One factor that is worthy of comment is that the cohorts have somewhat different median ages, and it is known that tumour molecular subtype has an incidence that depends on age. Since it is known that molecular subtypes can have different  $^1\text{H}$ -MRS, this is a potential confounder.<sup>31</sup> The SNR is not better in the 3 T compared with the 1.5 T cohort, as might have been expected, because a smaller voxel size is commonly used in the 3 T acquisition to reduce the partial volume effects even after increasing the number of excitations. In addition, the small number of ependymomas in the 3 T cohort leads to the balanced classification accuracy being a rather unstable measure of accuracy, with a large

reduction in its value if just one case is misclassified. Including supratentorial ependymomas to increase the number of these tumours also makes this class more diverse in biology and probably in metabolite profile. This combination of factors probably explains the poorer accuracy of classifications for the 3 T than the 1.5 T cohort.

Many metabolites might be highly interrelated through their biochemical pathways, but the relationships between metabolites have not been considered in the current study. Where significant correlation between metabolites is established, this could be used to reduce the number of metabolites required to achieve high classification accuracy. Future work to improve the accuracy of metabolite determination particularly for metabolites present at lower concentrations will be important.

## 5 | CONCLUSION

Metabolite selection provides an effective method of feature selection for childhood brain tumour classification from  $^1\text{H}$  single-voxel MRS, comparing favourably to the classic method of PCA. The technique has the advantage of identifying the key discriminatory metabolites, thereby bridging transparently from diagnosis based on single metabolites to machine learning, making it attractive for clinical and biomedical uses. Multi-class ROCs is the preferred implementation for metabolite selection in situations where there are more than two diagnoses that need to be discriminated, since it provides a single metric for each metabolite combined with high accuracy.

## ACKNOWLEDGMENTS

D.Z. was partly funded by a doctoral scholarship from the Help Harry Help Others Cancer Charity. A.C.P. is funded through an NIHR Research Professorship, NIHR-RP-R2-12-019. We would also like to acknowledge funding from the Cancer Research UK and EPSRC Cancer Imaging Programme at the Children's Cancer and Leukaemia Group (CCLG) in association with the MRC and Department of Health (England) (C7809/A10342), the Cancer Research UK and NIHR Experimental Cancer Medicine Centre Paediatric Network (C8232/A25261), the Children's Research Fund, Birmingham Women's and Children's Hospital Charities, The Children's Cancer and Leukaemia Group—Little Princess Trust (2017/15 and 2019/01), Children with Cancer (15/118), Action Medical Research (GN2181), The Brain Tumour Charity (GN2181) and Health Data Research UK (HDR UK). HDR UK is funded by the UK Medical Research Council, Engineering and Physical Sciences Research Council, Economic and Social Research Council, Department of Health and Social Care (England), Chief Scientist Office of the Scottish Government Health and Social Care Directorates, Health and Social Care Research and Development Division (Welsh Government), Public Health Agency (Northern Ireland), British Heart Foundation and Wellcome Trust.

## CONFLICTS OF INTEREST

The authors declare no potential conflicts of interest.

More details on the classification accuracy are presented in the Supporting Information.

## AUTHOR CONTRIBUTIONS

In this study, D.Z. and A.C.P. contributed to conception and design; H.E.L.R., N.P.D., M.W., L.J.A., S.A., D.R.G., T.J., P.S.M., L.M., D.M., S.B., B.P., T.N.A., Y.S. and A.C.P. contributed to functional imaging acquisition; J.T.G. and A.C.P. contributed to  $^1\text{H}$ -MRS review; D.Z., J.T.G. and A.C.P. contributed to  $^1\text{H}$ -MRS analysis and interpretation; D.Z. drafted the article; J.T.G. and A.C.P. revised the article critically for important intellectual content; H.E.L.R., M.W., V.S. and T.N.A. contributed to the final version of the article; all authors gave a final approval of the version to be published.

## DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

## ORCID

Dadi Zhao  <https://orcid.org/0000-0001-7727-3638>

James T. Grist  <https://orcid.org/0000-0001-7223-4031>

Heather E.L. Rose  <https://orcid.org/0000-0002-0346-1334>

Nigel P. Davies  <https://orcid.org/0000-0002-7978-5511>

Martin Wilson  <https://orcid.org/0000-0002-2089-3956>

Shivaram Avula  <https://orcid.org/0000-0003-1742-7206>

Tim Jaspan  <https://orcid.org/0000-0002-3644-9569>

Paul S. Morgan  <https://orcid.org/0000-0002-5870-1446>

Simon Bailey  <https://orcid.org/0000-0003-4763-4329>

Vijay Sawlani  <https://orcid.org/0000-0001-6779-1255>  
 Theodoros N. Arvanitis  <https://orcid.org/0000-0001-5473-135X>  
 Yu Sun  <https://orcid.org/0000-0001-6729-139X>  
 Andrew C. Peet  <https://orcid.org/0000-0002-4846-5152>

## REFERENCES

- Rickert CH, Paulus W. Epidemiology of central nervous system tumors in childhood and adolescence based on the new WHO classification. *Child's Nervous Syst.* 2001;17(9):503-511.
- Louis DN, Perry A, Reifenberger G, et al. The 2016 World Health Organization classification of tumours of the central nervous system: a summary. *Acta Neuropathol.* 2016;131(6):803-820.
- Fetit AE, Novak J, Peet AC, Arvanitis TN. Three-dimensional textural features of conventional MRI improve diagnostic classification of childhood brain tumours. *NMR Biomed.* 2015;28(9):1174-1184.
- Wilson M, Andronesi O, Barker PB, et al. Methodological consensus on clinical proton MRS of the brain: review and recommendations. *Magn Reson Med.* 2019;82(2):527-550.
- Kumar M, Arlauckas SP, Saksena S, et al. Magnetic resonance spectroscopy for detection of choline kinase inhibition in the treatment of brain tumours. *Mol Cancer Ther.* 2015;14(4):899-908.
- Kovanlikaya A, Panigrahy A, Krieger MD, et al. Untreated paediatric primitive neuroectodermal tumour in vivo: quantitation of taurine with MR spectroscopy. *Radiology.* 2005;236(3):1020-1025.
- Opstad KS, Bell BA, Griffiths JR, Howe FA. Taurine: a potential marker of apoptosis in gliomas. *Br J Cancer.* 2009;100(5):789-794.
- Davies NP, Wilson M, Natarajan K, et al. Non-invasive detection of glycine as a biomarker of malignancy in childhood brain tumours using in-vivo 1H MRS at 1.5 Tesla confirmed by ex-vivo high-resolution magic-angle spinning NMR. *NMR Biomed.* 2010;23(1):80-87.
- Jain M, Nilsson R, Sharma S, et al. Metabolite profiling identifies a key role for glycine in rapid cancer cell proliferation. *Science.* 2012;336(6084):1040-1044.
- Wilson M, Gill SK, MacPherson L, English M, Arvanitis TN, Peet AC. Non-invasive detection of glutamate predicts survival in paediatric medulloblastoma. *Clin Cancer Res.* 2014;20(17):4532-4539.
- Babourina-Brooks B, Kohe S, Gill SK, et al. Glycine: a non-invasive imaging biomarker to aid magnetic resonance spectroscopy in the prediction of survival in paediatric brain tumours. *Oncotarget.* 2018;9(27):18858-18868.
- Peet AC, Arvanitis TN, Leach MO, Waldman AD. Functional imaging in adult and paediatric brain tumours. *Nat Rev Clin Oncol.* 2012;9(12):700-711.
- Manias KA, Gill SK, Zarinabad N, et al. Evaluation of the added value of 1H-magnetic resonance spectroscopy for the diagnosis of pediatric brain lesions in clinical practice. *Neuro-Oncol Pract.* 2017;5(1):18-27.
- Manias KA, Gill SK, MacPherson L, Foster K, Oates A, Peet AC. Magnetic resonance imaging based functional imaging in paediatric oncology. *Eur J Cancer.* 2017;72:251-265.
- Manias KA, Gill SK, MacPherson L, et al. Diagnostic accuracy and added value of qualitative radiological review of 1H-magnetic resonance spectroscopy in evaluation of childhood brain tumors. *Neuro-Oncol Pract.* 2019;6(6):428-437.
- Vicente J, Fuster-Garcia E, Tortajada S, et al. Accurate classification of childhood brain tumours by in vivo 1H MRS—a multi-centre study. *Eur J Cancer.* 2013;49(3):658-667.
- Davies NP, Wilson M, Harris LM, et al. Identification and characterisation of childhood cerebellar tumours by in vivo proton MRS. *NMR Biomed.* 2008; 21(8):908-918.
- Jolliffe IT, Cadima J. Principal component analysis: a review and recent developments. *Philos Trans R Soc A.* 2016;374(2065):20150202.
- Orphanidou-Vlachou E, Kohe SE, Brundler MA, et al. Metabolite levels in paediatric brain tumours correlate with histological features. *Pathobiology.* 2018;85(3):157-168.
- Wilson M, Reynolds G, Kauppinen RA, Arvanitis TN, Peet AC. A constrained least-squares approach to the automated quantitation of in vivo 1H magnetic resonance spectroscopy data. *Magn Reson Med.* 2011;65(1):1-12.
- Hand DJ, Till RJ. A simple generalisation of the area under the ROC curve for multiple class classification problems. *Mach Learn.* 2001;45(2):171-186.
- Zarinabad N, Abernethy LJ, Avula S, et al. Application of pattern recognition techniques for classification of paediatric brain tumours by in vivo 3T 1H-MR spectroscopy—a multi-centre study. *Magn Reson Med.* 2018;79(4):1-8.
- Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. *J Artif Intell Res.* 2002;16:321-357.
- Wong T-T. Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation. *Pattern Recog.* 2015;48(9):2839-2846.
- Srinivasan R, Vigneron D, Sailasuta N, Hurd R, Nelson S. A comparative study of myo-inositol quantification using LCMoDel at 1.5 T and 3.0 T with 3D 1H proton spectroscopic imaging of the human brain. *Magn Reson Imaging.* 2004;22(4):523-528.
- Bennett CD, Kohe SE, Gill SK, et al. Tissue metabolite profiles for the characterisation of paediatric cerebellar tumours. *Sci Rep.* 2018;8(1):11992.
- Harris LM, Davies NP, Macpherson L, et al. Magnetic resonance spectroscopy in the assessment of pilocytic astrocytomas. *Eur J Cancer.* 2008;44(17): 2640-2647.
- Orphanidou-Vlachou E, Auer D, Brundler MA, et al. 1H magnetic resonance spectroscopy in the diagnosis of paediatric low grade brain tumours. *Eur J Radiol.* 2013;82(6):e295-301.
- Carlin D, Babourina-Brooks B, Davies NP, Wilson M, Peet AC. Variation of T2 relaxation times in pediatric brain tumors and their effect on metabolite quantification. *J Magn Reson Imaging.* 2019;49(1):195-203.
- Obuchowski NA. Receiver operating characteristic curves and their use in radiology. *Radiology.* 2003;229(1):3-8.
- Blüml S, Margol AS, Sposto R, et al. Molecular subgroups of medulloblastoma identification using noninvasive magnetic resonance spectroscopy. *Neuro-Oncology.* 2016;18(1):126-131.

**SUPPORTING INFORMATION**

Additional supporting information may be found in the online version of the article at the publisher's website.

**How to cite this article:** Zhao D, Grist JT, Rose HEL, et al. Metabolite selection for machine learning in childhood brain tumour classification. *NMR in Biomedicine*. 2022;35(6):e4673. doi:[10.1002/nbm.4673](https://doi.org/10.1002/nbm.4673)