

Applying machine learning classifiers to automate quality assessment of paediatric dynamic susceptibility contrast (DSC-) MRI data

Powell, Stephen J; Withey, Stephanie B; Sun, Yu; Grist, James T; Novak, Jan; Macpherson, Lesley; Abernethy, Laurence; Pizer, Barry; Grundy, Richard; Morgan, Paul S; Jaspan, Tim; Bailey, Simon; Mitra, Dipayan; Auer, Dorothee P; Avula, Shivaram; Arvanitis, Theodoros N; Peet, Andrew

DOI:
[10.1259/bjr.20201465](https://doi.org/10.1259/bjr.20201465)

License:
Creative Commons: Attribution (CC BY)

Document Version
Publisher's PDF, also known as Version of record

Citation for published version (Harvard):
Powell, SJ, Withey, SB, Sun, Y, Grist, JT, Novak, J, Macpherson, L, Abernethy, L, Pizer, B, Grundy, R, Morgan, PS, Jaspan, T, Bailey, S, Mitra, D, Auer, DP, Avula, S, Arvanitis, TN & Peet, A 2023, 'Applying machine learning classifiers to automate quality assessment of paediatric dynamic susceptibility contrast (DSC-) MRI data', *British Journal of Radiology*, vol. 96, no. 1145, 20201465. <https://doi.org/10.1259/bjr.20201465>

[Link to publication on Research at Birmingham portal](#)

General rights

Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

- Users may freely distribute the URL that is used to identify this publication.
- Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
- User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
- Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

Take down policy

While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact UBIRA@lists.bham.ac.uk providing details and we will remove access to the work immediately and investigate.

Download date: 29. Apr. 2024

Received:
21 December 2020Accepted:
24 January 2023Published online:
16 February 2023

© 2023 The Authors. Published by the British Institute of Radiology under the terms of the Creative Commons Attribution 4.0 Unported License <http://creativecommons.org/licenses/by/4.0/>, which permits unrestricted use, distribution and reproduction in any medium, provided the original author and source are credited.

Cite this article as:

Powell SJ, Withey SB, Sun Y, Grist JT, Novak J, MacPherson L, et al. Applying machine learning classifiers to automate quality assessment of paediatric dynamic susceptibility contrast (DSC-) MRI data. *Br J Radiol* (2023) 10.1259/bjr.20201465.

FULL PAPER

Applying machine learning classifiers to automate quality assessment of paediatric dynamic susceptibility contrast (DSC-) MRI data

^{1,2}STEPHEN J. POWELL, ^{2,3,4}STEPHANIE B. WITHEY, ^{2,5}YU SUN, ²JAMES T. GRIST, ^{2,3,6}JAN NOVAK, ⁷LESLEY MACPHERSON, ⁸LAURENCE ABERNETHY, ⁹BARRY PIZER, ¹⁰RICHARD GRUNDY, ^{10,11,12}PAUL S. MORGAN, ^{10,13}TIM JASPAN, ¹⁴SIMON BAILEY, ¹⁵DIPAYAN MITRA, ¹⁶DOROTHEE P. AUER, ⁸SHIVARAM AVULA, ^{2,3,17}THEODOROS N. ARVANITIS and ^{2,3}ANDREW PEET

¹Physical Sciences for Health CDT, University of Birmingham, Birmingham, United Kingdom

²Institute of Cancer and Genomic Sciences, University of Birmingham, Birmingham, United Kingdom

³Department of Oncology, Birmingham Children's Hospital, Birmingham, United Kingdom

⁴RRPPS, University Hospitals Birmingham NHS Foundation Trust, Birmingham, United Kingdom

⁵School of Biological Sciences and Medical Engineering, Southeast University, Nanjing, China

⁶Department of Psychology, Aston Brain Centre, School of Life and Health Sciences, Aston University, Birmingham, United Kingdom

⁷Radiology, Birmingham Children's Hospital, Birmingham, United Kingdom

⁸Radiology, Alder Hey Children's NHS Foundation Trust, Liverpool, United Kingdom

⁹Oncology, Alder Hey Children's NHS Foundation Trust, Liverpool, United Kingdom

¹⁰The Children's Brain Tumour Research Centre, University of Nottingham, Nottingham, United Kingdom

¹¹Medical Physics, Nottingham University Hospitals, Nottingham, United Kingdom

¹²NIHR Nottingham Biomedical Research Centre, Nottingham, United Kingdom

¹³Radiology, Nottingham University Hospitals, Nottingham, United Kingdom

¹⁴Sir James Spence Institute of Child Health, Royal Victoria Infirmary, Newcastle upon Tyne, United Kingdom

¹⁵Neuroradiology, The Newcastle upon Tyne Hospitals NHS Foundation Trust, Newcastle upon Tyne, United Kingdom

¹⁶Sir Peter Mansfield Imaging Centre, University of Nottingham, Nottingham, United Kingdom

¹⁷Institute of Digital Healthcare, WMG, University of Warwick, Coventry, United Kingdom

Address correspondence to: Professor Andrew Peet

E-mail: a.peet@bham.ac.uk

Objective: Investigate the performance of qualitative review (QR) for assessing dynamic susceptibility contrast (DSC-) MRI data quality in paediatric normal brain and develop an automated alternative to QR.

Methods: 1027 signal-time courses were assessed by Reviewer 1 using QR. 243 were additionally assessed by Reviewer 2 and % disagreements and Cohen's κ (κ) were calculated. The signal drop-to-noise ratio (SDNR), root mean square error (RMSE), full width half maximum (FWHM) and percentage signal recovery (PSR) were calculated for the 1027 signal-time courses. Data quality thresholds for each measure were determined using QR results. The measures and QR results trained machine learning classifiers. Sensitivity, specificity, precision, classification error and area under the curve from a receiver operating characteristic curve were calculated for each threshold and classifier.

Results: Comparing reviewers gave 7% disagreements and $\kappa = 0.83$. Data quality thresholds of: 7.6 for SDNR; 0.019 for RMSE; 3s and 19s for FWHM; and 42.9 and 130.4% for PSR were produced. SDNR gave the best sensitivity, specificity, precision, classification error and area under the curve values of 0.86, 0.86, 0.93, 14.2% and 0.83. Random forest was the best machine learning classifier, giving sensitivity, specificity, precision, classification error and area under the curve of 0.94, 0.83, 0.93, 9.3% and 0.89.

Conclusion: The reviewers showed good agreement. Machine learning classifiers trained on signal-time course measures and QR can assess quality. Combining multiple measures reduces misclassification.

Advances in knowledge: A new automated quality control method was developed, which trained machine learning classifiers using QR results.

INTRODUCTION

Dynamic susceptibility contrast (DSC-) MRI provides estimates of perfusion in the brain,¹ by imaging the passage of

a gadolinium-based contrast agent using a dynamic T_2 or T_2^* weighted imaging sequence.² The contrast agent causes local changes in T_2 and T_2^* , which dynamically alter the MR

signal intensity.³ Analysis of the resulting signal–time courses, associated with each pixel, can produce estimates of cerebral blood volume (CBV), cerebral blood flow (CBF) and vascular mean transit time (MTT).⁴ As well as DSC-MRI, perfusion can also be measured using MRI with dynamic contrast-enhanced (DCE-) MRI, arterial spin labelling (ASL) and intravoxel incoherent motion (IVIM). However, DSC-MRI offers better signal-to-noise ratio (SNR) and contrast-to-noise ratio, and a faster acquisition time.⁵

Measurement of perfusion can be used to indicate health in a range of diseases. In paediatrics, it is used to assess brain tumours, which are the leading cause of cancer-related mortality in children,⁶ as well as diseases which affect neurovasculature. Most paediatric brain tumour patients have a gadolinium injection to allow for post-contrast T_1 weighted imaging so the DSC-MRI acquisition can be carried out during this injection providing information that would otherwise not be available.⁷ CBV and CBF values from DSC-MRI acquisitions in paediatric patients have been used to predict long-term survival⁸ and have been shown to correlate with tumour grade.^{9,10} These applications require accurate CBV and CBF values, therefore it is important to ensure that the DSC-MRI signal–time courses they are estimated from are of good quality.¹¹

DSC-MRI is prone to motion and susceptibility artefacts, which degrade the quality of acquired data.¹¹ The scanner and acquisition protocol for DSC-MRI also commonly varies from centre-to-centre, which affects the signal–time courses produced^{12,13} and the SNR of the CBV maps,¹⁴ limiting the clinical applicability of the technique.¹⁵ For example, the field strength of the scanner and acquisition parameters such as repetition time (TR), echo time (TE), voxel volume and flip angle may vary between centres. These factors affect the SNR of the acquired data, whilst TR dictates the temporal resolution.¹⁶ In brain tumour patients, breakdown of the blood–brain barrier (BBB) can lead to contrast agent extravasation, where the contrast agent leaks into the extravascular extracellular space (EES).¹⁷ Contrast extravasation can lead to T_1 weighted contamination of the signal–time courses and underestimation of the CBV values, or T_2^* weighted effects leading to overestimation of CBV values.¹⁷ These contamination effects can be reduced either by administering a pre-bolus of contrast agent,¹⁸ or using a low flip angle during the acquisition, to reduce the T_1 weighting of the DSC-MRI sequence.¹¹ Recent research has shown that the application of leakage correction is essential when using a low flip angle, single-bolus protocol.^{7,19,20}

Currently, the ASFNIR recommendation for quality control (QC) of DSC-MRI data is to assess signal–time courses by eye, using qualitative review (QR). This involves assessing signal–time courses for the presence of artefacts, including magnetic susceptibility (the response of a material to a magnetic field, which can result in signal loss²¹ and motion); for appropriate signal drop indicating the quality of bolus administration; and for noise spikes in the signal–time curve, suggesting that any such time points should be removed.¹¹ There can be discordance between reviewers and one DSC data set contains thousands of signal–time courses, so it is not practical to assess the quality of

all signal–time courses manually. In practice this means that a subset of the signal–time courses is used to assess the quality of a whole data set. An automated process based on QR, which could be applied to assess signal–time course quality on a voxel-wise basis, which could be used to provide an assessment of the overall quality of a data set, is desirable.

Previous work has shown that it is possible to define statistical thresholds and apply these to quantitative measures calculated from DSC-MRI signal–time courses to assess data quality.²² Machine learning (ML) classification can be used to train models to make predictions based on features extracted from a data set.²³ This has plenty of applications in medical imaging. For example, it has been used in the pneumonia detection,^{24,25} detection and classification of Covid-19,²⁶ diagnosing colorectal cancer,²⁷ and assisting in planning rehabilitation care for stroke patients.²⁸ ML has also been applied to DSC-MRI data for several applications, but so far it has not been used for assessing data quality. For example, it has been used in place of standard analysis techniques for DSC-MRI to provide a quicker and more robust method to estimate CBF values from raw signal–time course data,²⁹ predict survival in glioma patients,³⁰ and classify tumour type.^{31–36} Therefore, ML could be applied to features extracted from DSC-MRI signal–time courses to determine data quality. Any new method for assessing data quality should be established in normal brain before it is applied to diseased tissues. Undertaking such a study is a challenge in children due to ethical constraints but an appropriate alternative is to use paediatric patients undergoing DSC-MRI scans for brain tumours, and selecting signal–time courses from slices of brain which do not contain tumour.

The objectives of this paediatric study are: to assess the discordance in QR between two reviewers, to use this QR to determine thresholds of quantitative measures of data quality and to investigate whether QR and quantitative thresholds could be used to develop an automated QC process for assessing overall data quality of a paediatric data set.

METHODS

Patient data

For this study, a data set containing 25 paediatric patients, acquired at 4 UK centres was used. The data were gathered from the Children's Cancer and Leukaemia Group (CCLG) functional imaging of tumours database.³⁷ 23 of the data sets were acquired pre-diagnosis, and 2 of the data sets were acquired post-diagnosis. One of the post-diagnosis patients underwent a biopsy and chemotherapy, and the other underwent a surgical resection. The acquisition protocols used are summarised in Table 1. The patient data are from an imaging study entitled "CNS 2004 10 Functional Imaging of Tumours" (NRES REC ref: 04/MRE04/41). This is a multicentre paediatric study with ref: RG_09–028 and ethics ref: ERN_11–1170. Informed parental consent was obtained for all patients included in the study.

Figure 1 summarises the processes involved in developing an automated QC aid. Details of each step are given below.

Table 1. A summary of each of the acquisition protocols used, the number of patients who were recorded using that protocol and the number of signal-time courses used

Centre	No. of patients	No. of time courses	Field strength (T)	Pre-bolus	Sequence	Flip angle (°)	TE (ms)	TR (ms)	Voxel size (mm)
1	3	139	3	Yes	GE-EPI	20	40	1829–4865	2.5 × 2.5 × 3.5
1	5	193	1.5	Yes	GE-EPI	20	40	1490–1643	2.4 × 2.4 × 5
1	3	112	3	No	sPRESTO	7	22	15	3.4 × 3.4 × 3.5
2	5	177	3	Yes	GE-EPI	75	40	1335–2343	1.75 × 1.75 × 4
3	3	166	3	No	GE-EPI	45	29	1570	3.4 × 3.4 × 3.5
4	3	112	3	Yes	GE-EPI	20	40	1865	2.5 × 2.5 × 3.5
4	2	96	1.5	No	sPRESTO	7	25	17	3.4 × 3.4 × 3.5
4	1	16	3	No	sPRESTO	7	24	16	1.8 × 1.8 × 3.5

In the Sequence column, GE-EPI = Gradient Echo - Echo Planar Imaging, and sPRESTO = Sensitivity Encoded (SENSE) Principles of Echo-Shifting with a Train of Observations.³⁸

QR of patient data

QR of 1027 signal-time courses, extracted from 25 patients, was performed. A large number of patients were used to ensure a range of acquisition protocols and artefacts were included. Artefacts were those observed in normal clinical practice when scanning patients. Table 1 summarises how many signal-time courses came from each acquisition protocol. Signal-time courses were randomly selected from pre-defined regions within each patient, which included: grey matter (GM), white matter (WM), the edge of the brain, the edge of the ventricles and the cerebellum. All signal-time courses were selected from slices which did not contain any tumour, by selecting supratentorial signal-time courses from patients with infratentorial tumours and infratentorial signal-time courses from patients with supratentorial tumours. Tumour diagnosis information was obtained from the CCLG database. All signal-time courses were assessed using QR by Author 1 (PhD student with 3 years' experience), and a randomly selected subset of 243 signal-time courses were additionally assessed by Author 2 (Clinical scientist with 8 years' experience). QR to assess data quality was carried out using the guidance from the ASFNR recommendations.¹¹ This involved assessing whether a clear signal drop was present and the level of noise within the baseline and the rest of the signal. Signal-time courses were then given a score of 1 (accepted) or 0 (rejected) based on this assessment. For the subset of signal-time courses reviewed by two reviewers, the scores from Author 2 were considered to be the ground truth and were used for the determination of thresholds and in the training of ML classifiers.

The percentage disagreement between the two reviewers and the Cohen's κ for interrater reliability were calculated for the subset of 243 signal-time courses assessed by both reviewers. All

statistical analysis was carried out in R (R Foundation for Statistical Computing, Vienna, Austria, v. 3.5.0).

Calculating the quantitative measures of quality

Signal drop-to-noise ratio (SDNR), root mean square error (RMSE), full width half maximum (FWHM) and percentage signal recovery (PSR) were used as quantitative measures of signal-time course quality. These were calculated for each of the 1027 signal-time courses which had previously undergone QR. SDNR was calculated using equation 1, with the signal drop defined as the difference between the mean baseline and mean of the first pass minima and the two adjacent dynamics. This is a similar measure to the contrast-to-noise ratio applied in work by Digernes et al, except it is calculated from the signal-time course instead of the relaxation rate curve.³⁹

$$SDNR = \frac{\text{SignalDrop}}{\text{StandardDeviation in Baseline}} \quad (1)$$

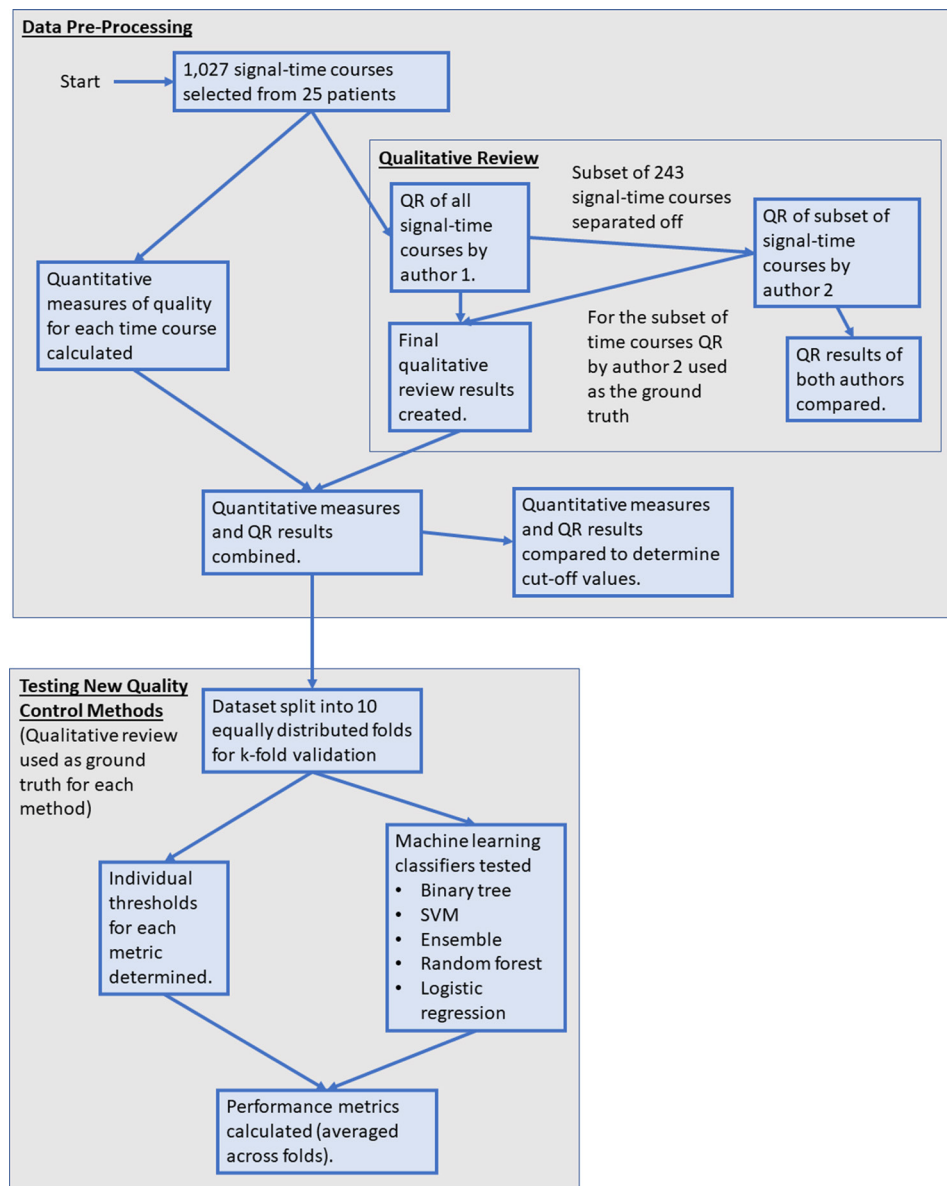
RMSE was calculated by fitting a version of the simplified γ variate function,⁴⁰ shown in equation 2, to the first pass of the signal-time course.

$$y(t) = c - Kt^\alpha e^{-\frac{t}{\beta}} \quad (2)$$

Where $y(t)$ is the fit, t is the time, c is the average baseline signal, and α , β and K are shape coefficients. The RMSE value from this fit was normalised to the area of the first pass.

The FWHM was calculated as the width of the first pass (in seconds) at half the signal drop.

Figure 1. A flowchart summarising the various stages of the work presented within this paper. QR, qualitative review; SVM, support vector machine.



The PSR was calculated from equation 3, with T_2^* recovery defined as the difference between the mean post-bolus signal and the mean of the first pass minima and the two adjacent dynamics.¹¹

$$PSR = \frac{T_2^* \text{Recovery}}{\text{SignalDrop}} \times 100 \quad (3)$$

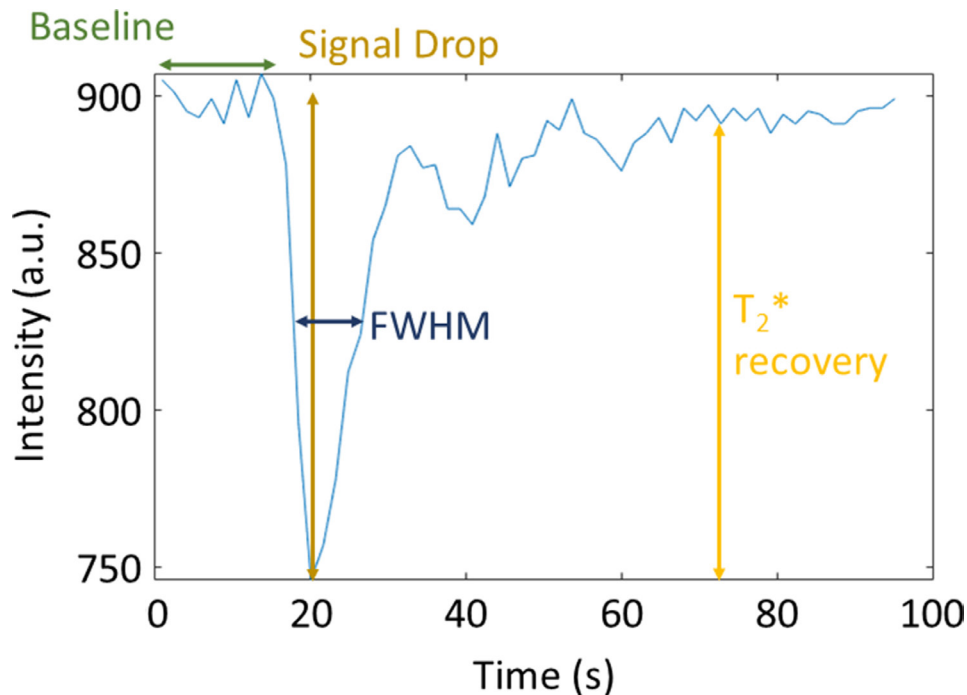
To calculate these measures, it was necessary to define the dynamics where the baseline ended, and the post-bolus started. The baseline end was determined by calculating the moving mean (with sliding window of three) and cumulative mean of the signal-time course, starting from the first dynamic, and finding the dynamic where the means diverged. The start of the post-bolus was determined using the same process but starting from the last timepoint. The first pass was defined as the region

between the end of the baseline and the start of the post-bolus. Figure 2 shows an example signal-time course and the features used to calculate the quantitative measures.

Thresholds from QR

Quality thresholds for SDNR, RMSE, FWHM, and PSR, were determined using the QR results from the 1027 signal-time courses that underwent QR (Figure 1). Thresholds were determined using k-fold cross-validation (CV), with $k = 10$. Data are separated into k equally sized folds, with $(k-1)$ folds used as training data, and the remaining fold used as testing data, from which the performance metrics are calculated. This process is repeated until all folds have been used as testing data.⁴¹ The separation of signal-time courses into folds was stratified to ensure an even distribution of accepted and rejected signal-time courses

Figure 2. Example signal-time course showing which features of the signal-time course are used to calculate the quantitative measures. FWHM, full width half maximum.



in each fold. The centre of data set the signal-time courses came from was not considered when separating into folds.

For each fold, threshold values were determined from the training data. Sensitivity, specificity, precision, classification error and area under curve (AUC) from a receiver operator curve (ROC) were calculated as performance metrics, by applying the thresholds to the testing data. Mean thresholds and performance metrics were calculated by averaging across the folds.

SDNR and RMSE quality thresholds were determined using sensitivity vs specificity plots. For each fold, the SDNR threshold was varied over each SDNR value within the training data, and the sensitivity and specificity were calculated from applying the threshold to the training data and comparing to the QR results. The optimal threshold was the value where sensitivity equalled specificity. This process was repeated for the RMSE values.

Upper and lower thresholds of quality were determined for FWHM and PSR, respectively. For each fold, the parameter values from the training data were ordered in ascending value. The signal-time courses with the smallest and largest FWHM or PSR values, respectively, that passed QR were identified. These values were used as thresholds, with any signal-time course with an FWHM or PSR between the two thresholds classed as good quality.

Combining quantitative measures using ML

ML classification was carried out using the ML toolbox in Matlab (The MathWorks, MA, 2019a).⁴² Classification was carried out using the data set of 1027 signal-time courses used for determining thresholds from QR. SDNR, RMSE, FWHM and PSR

values were used as predictors for classifier training and the QR scores (1 = passed QR, 0 = failed QR) used as the target outputs. Hyperparameter optimisation was applied for each classifier and k-fold CV with $k = 10$ was used. As previously the k-fold validation was stratified to ensure that there an even distribution of accepted and rejected signal-time courses in each fold, but the centre of data set the signal-time course came from was not taken into consideration, as the aim of this work is to apply the final classifier to a wide range of patient data. The classifiers used were binary tree, support vector machine (SVM), ensemble, random forest, and logistic regression. These classifiers were selected to ensure that a wide range of classification methods were applied to the data. This will help to ensure that the optimal ML classifier is chosen. The average sensitivity, specificity, precision, classification error and AUC were calculated for each classifier. Further details on the ML, including the hyperparameter optimisation and the results of the hyperparameter optimisation, can be found in the appendix.

Application to patient data

Each of the thresholds of the quantitative measures of quality and the best performing ML classifier were applied to signal-time courses obtained from one slice of patient data acquired using the acquisition protocol described in row 3 of Table 1. A quality map was created for each method, showing which voxels had passed QC and which had failed.

RESULTS

QR of patient data shows good agreement between reviewers, and that there is a region of

Table 2. A summary of the differences between reviewers, in terms of percentage difference and Cohen's κ , across all time courses, all time courses recorded at 1.5 T, and all time courses recorded at 3 T

Group	No. of time courses	Percentage difference (%)	Cohen's κ
All signal-time courses	243	6.58	0.84
1.5 T signal-time courses	81	12.35	0.73
3 T signal-time courses	162	3.70	0.79

uncertainty where it is difficult to classify signal-time courses

Table 2 splits the signal-time courses assessed by two reviewers into three groups (all signal-time courses, all 1.5 T signal-time courses, and all 3 T signal-time courses) and summarises the percentage disagreements and Cohen's κ for each group. Across the entire subset, the signal-time courses where there were disagreements between reviewers had median SDNR of 5.4 (range 3.3–56.4), median RMSE of 0.020 (range 0.005–0.058), median PSR of 84.5% (range 52.5–107.1%), and median FWHM of 7 s (range 4–15 s).

Figure 3 shows some examples of signal-time courses that failed QR for different reasons: 3(a) failed due to a small signal drop in comparison to the noise in the baseline, leading to a low SDNR; 3(b) failed due to a noisy first pass, leading to a large RMSE value; 3(c) failed due to a very narrow first pass, which lead to a very small FWHM value; 3(d) failed due to a low T_2^* recovery, resulting in a low PSR value. Below an SDNR value of

2.8, no signal-time courses passed QR, whilst above SDNR of 56.4, no signal-time courses failed QR. Above an RMSE value of 0.0846, no signal-time courses passed QR, whilst below an RMSE value of 0.0055, no signal-time courses failed QR. No signal-time courses with FWHM values less than 3 s or greater than 19 s passed QR. No signal-time courses with PSR values less than 43% or greater than 131% passed QR.

SDNR is the most important quantitative measure, but other measures are needed

Figure 4 shows example sensitivity vs specificity plots for one of the folds. Averaging across all folds produced an SDNR threshold of 7.6, an RMSE threshold of 0.019, FWHM lower and upper thresholds of 3 s and 19 s, and PSR lower and upper thresholds of 42.9 and 130.4%. The average performance measures are summarised in Table 3. Figure 5 shows example signal-time courses where there were disagreements between the QR results and the SDNR threshold for acceptance of quality. Out of the three signal-time courses that passed the SDNR threshold but

Figure 3. An example of signal-time course that failed QR for differing reasons. (a) has a low SDNR value of 2.7, (b) has a large RMSE value of 0.129, (c) has a small FWHM value of 1.8s (d) has a low PSR value of 36.9%. FWHM, full width half maximum; PSR, percentage signal recovery; RMSE, root mean square error; SDNR, signal drop-to-noise ratio.

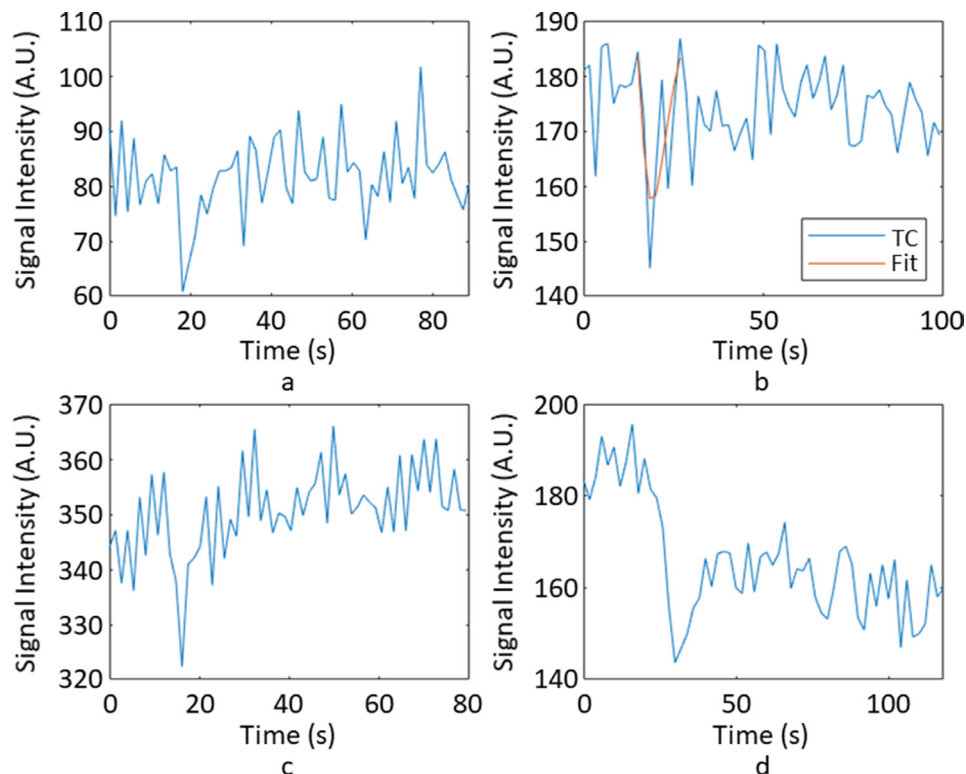
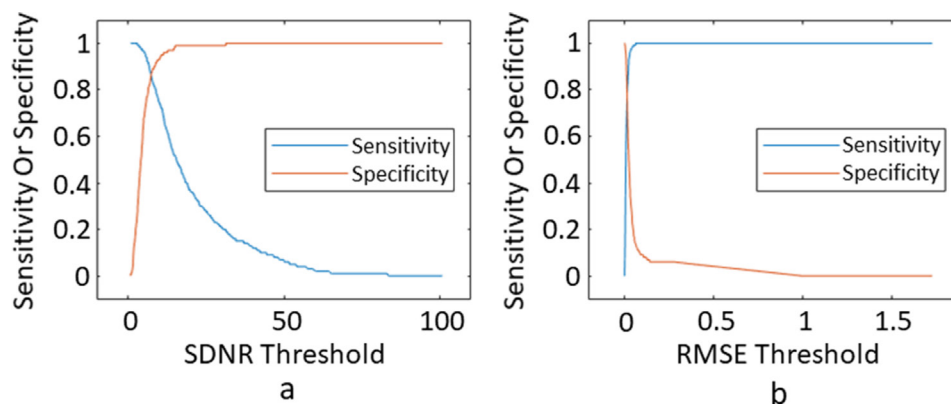


Figure 4. Sensitivity vs. specificity plots for (a) SDNR, (b) RMSE. RMSE, root mean square error; SDNR, signal drop-to-noise ratio.



failed QR, all three passed the FWHM and PSR thresholds, and one passed the RMSE threshold. All three signal–time courses failed QR because of issues with the post-bolus signal, which was not picked up by SDNR. Out of the three signal–time courses that passed QR but failed the SDNR threshold, one passed the RMSE threshold and two passed the FWHM and PSR thresholds.

Combining quantitative measures to assess data quality reduces classification error and ML classifiers offer an automated method to do this. Table 4 summarises the average performance measures for each of the ML classifiers. The classifier with the lowest classification error was the random forest, producing sensitivity, specificity, precision, classification error and AUC of 0.94, 0.83, 0.93, 9.3% and 0.89, respectively. Figure 6 shows an example of the confusion matrix and the ROC curve from the best performing fold of the best performing classifier. Figure 7 shows examples of the disagreements between the QR and the ML results. Details of the hyperparameter optimisation and its results for the best performing classifier can be found in the appendix.

When applied to one slice of patient data ML passed more signal–time courses than the SDNR threshold

Figure 8 shows the quality maps produced by applying the thresholds as obtained from the QR of each of the described metrics and the random forest ML classifier, respectively, to one slice of patient data. Blue pixels represent signal–time courses that passed the respective QC method, whilst orange pixels

represent those that failed. Table 5 summarises the percentage of signal–time courses that passed each method.

DISCUSSION

Our study shows that although QR can be used to assess data quality, there are a range of signal–time courses which are difficult to classify. Automated quality control methods using simple metrics can be developed using the results of QR. Combining multiple metrics using ML results in fewer signal–time course misclassifications than using individual metrics. However, selecting a set of metrics to fully describe a signal–time course and all of its potential artefacts is challenging. We applied the automated QC methods to paediatric data in this work, however they are also applicable to adult data.

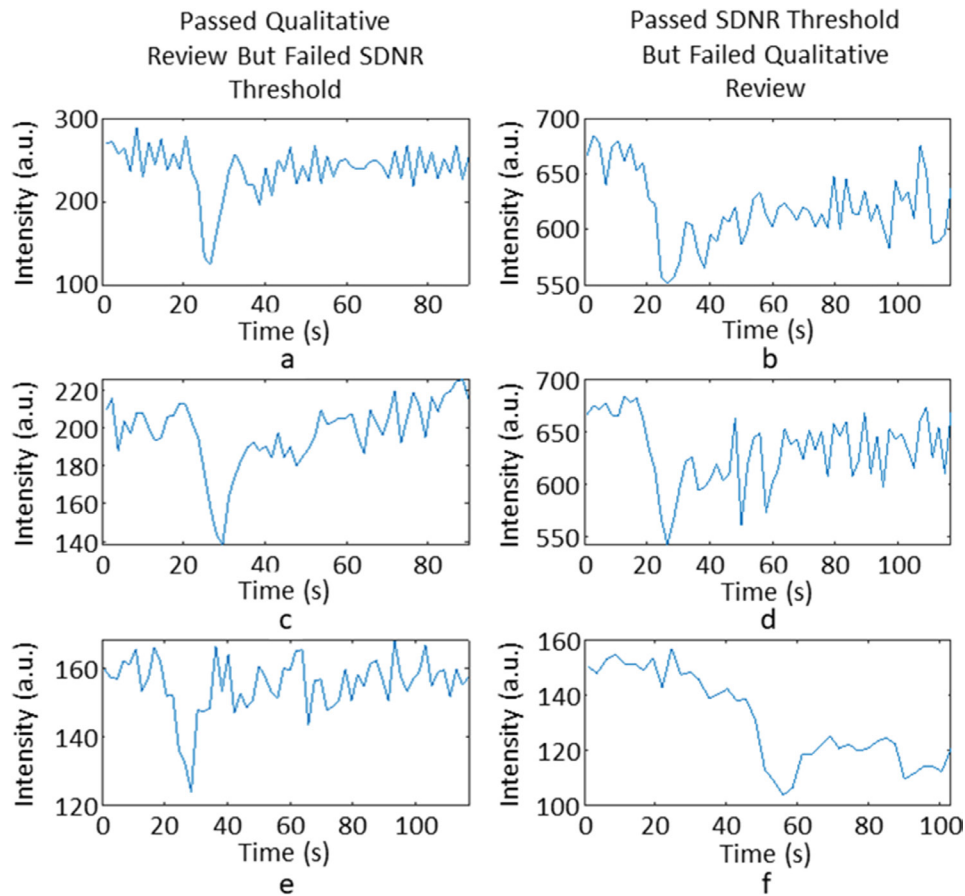
The signal–time courses assessed by two reviewers show a low discordance between reviewers, due to a low percentage of disagreements and a Cohen's κ value of 0.83, which shows excellent agreement.⁴³ The two reviewers found it harder to agree on whether to pass signal–time curves from data sets acquired at 1.5 T, due to their reduced SDNR. For the entire subset, the ranges of SDNR and RMSE values for signal–time courses where there were disagreements between the reviewers and for signal–time courses that were within the region of uncertainty are both large. This shows that a signal–time course with a large SDNR is not guaranteed to be good quality, as other factors may also affect quality, including, *e.g.* a distorted first pass.

Table 3. Summary of the average threshold values, sensitivity, specificity, precision, classification errors and AUCs for each of the quantitative measures

Quantitative measure	Threshold(s)	Sensitivity	Specificity	Precision	Classification error (%)	AUC
SDNR	7.6	0.86	0.86	0.93	14.2	0.83
RMSE	0.019	0.79	0.79	0.89	21.8	0.75
FWHM (s)	3, 19	1.00	0.18	0.74	24.8	0.85
PSR (%)	42.9, 130.4	1.00	0.12	0.73	26.4	0.84

AUC, area under the curve; FWHM, full width half maximum; PSR, percentage signal recovery; RMSE, root mean square error; SDNR, signal drop-to-noise ratio.

Figure 5. A demonstration of the disagreements between the QR results and the SDNR threshold. The left column (a, c, e) shows signal-time courses that passed QR but failed the SDNR threshold. The right column (b, d, f) shows signal-time courses that passed the SDNR threshold but failed QR. QR, qualitative review; SDNR, signal drop-to-noise ratio.



The automated QC methods presented in this work assess data quality on a voxel-wise basis, which allows for more data to be assessed than in QR. A typical DSC-MRI data set will have tens of thousands of signal-time courses. Assessing all of them by eye is not possible, so QR generally involves assessing a small subset of the whole data set. Our automated QC methods are ML classifiers, trained using signal-time courses from “normal brain”, which could lead to the exclusion of pathology related low perfusion signal-time courses, which may be clinically useful. Large amounts of WM, which is suggested by consensus guidelines as the tissue to use for normalising rCBV values,¹⁹ may also be excluded as it is less perfused than GM. A better approach

may be to use voxel-wise QC to give an overall assessment of data quality, which could then be used to decide whether a data set is of sufficient overall quality to be included in a study. An alternative to this would be to assess the quality of an average signal-time course from a data set. This would be quicker than a voxel-wise analysis but could produce misleading results—an average signal-time course will “average out” noise and regions of artefacts from the data.

The SDNR threshold defined the minimum SDNR for data to be accepted and was the best-performing individual measure, giving the most similar results to QR. This is expected as SDNR defines

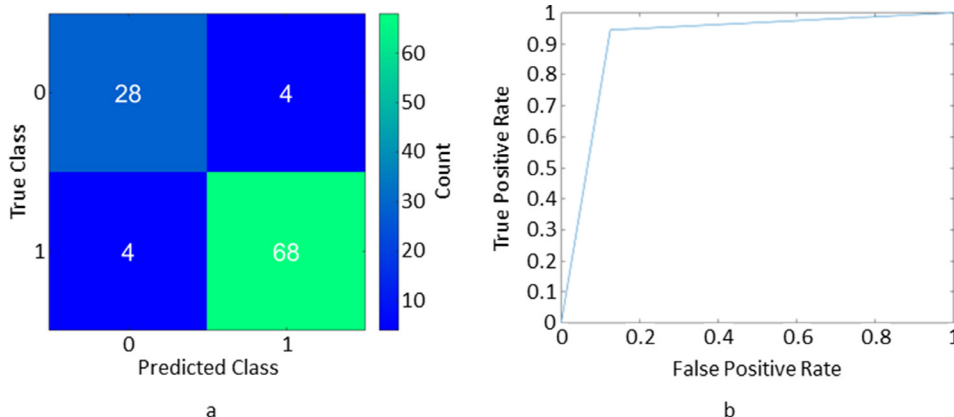
Table 4. Summary of the performance metrics for each machine learning classifier

Classifier	Sensitivity	Specificity	Precision	Classification error (%)	AUC
Binary Tree	0.93	0.80	0.92	11.4	0.87
SVM	0.93	0.84	0.93	9.5	0.89
Ensemble ^a	0.94	0.82	0.92	9.5	0.89
Random forest	0.94	0.83	0.93	9.3	0.89
Logistic regression	0.93	0.83	0.92	10.3	0.88

AUC, area under the curve; SVM, support vector machine.

^aThe Bag method was selected the by the hyperparameter optimisation.

Figure 6. A confusion matrix (a) and an ROC curve (b) from the best performing fold of the best performing classifier (Random forest). The ROC curve gave an AUC value of 0.90. AUC, area under the curve; ROC, receiver operating curve.



how visible the signal drop is, which is a key part of assessing data quality by QR.¹¹ However, SDNR is reduced in low perfused tissues, such as white matter or some low-grade tumours, and this may lead to the exclusion of some of these signal-time courses, discarding clinically useful information. The RMSE threshold gave poorer performance across all the performance measures compared to SDNR. The FWHM and PSR thresholds

give AUC values comparable to the other quantitative measures and a similar classification error to the RMSE threshold. Both resulted in very good sensitivity but poor specificity.

Multiple factors affect DSC-MRI data quality, and a single measure cannot cover them all. Figure 5 illustrates the difficulties of trying to classify signal-time course quality purely on SDNR.

Figure 7. An example of some of the agreements and disagreements between the ML results and QR: (a, c) show signal-time courses that failed ML but passed QR, (b, d) show signal-time courses that passed ML but failed QR. ML, machine learning; QR, qualitative review.

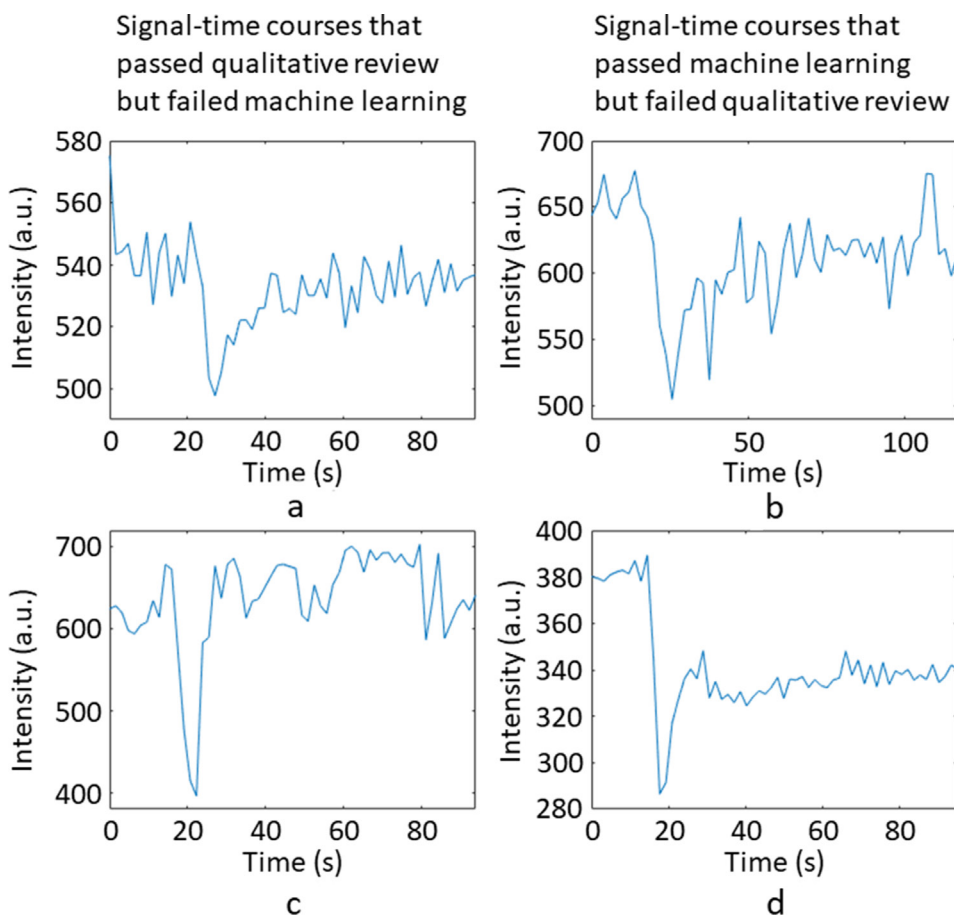
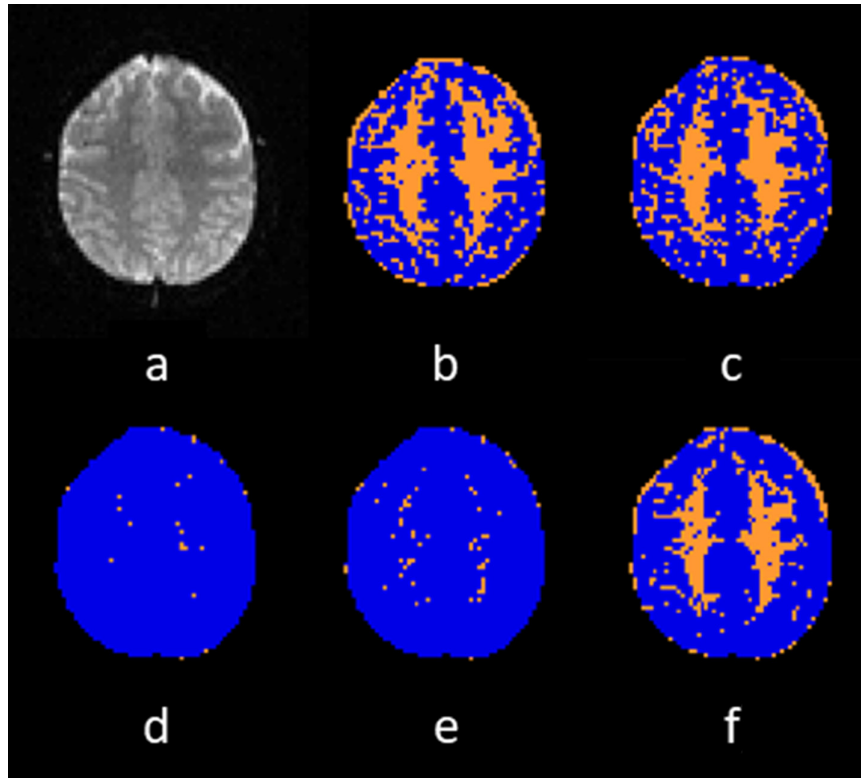


Figure 8. (a) An axial slice from a patient data set recorded at 1.5 T, and the resulting quality maps from applying (b) the SDNR threshold, (c) RMSE threshold, (d) PSR threshold, (e) the FWHM threshold, and (f) the ML classifier, as QC methods. Blue pixels represent time courses that passed QC and orange pixels represent time courses that failed QC. AUC, area under the curve; FWHM, full width half maximum; ML, machine learning; PSR, percentage signal recovery; RMSE, root mean square error; QC, quality control; SDNR, signal drop-to-noise ratio.



Difficulty in defining a single threshold for each quantitative measure shows the need for combined measures to assess data quality. This agrees with work by Akella *et al*²² where multiple quantitative measures including, the failure rate of fitting a gamma-variate to the first pass, mean FWHM, and mean PSR were calculated from the signal–time courses in each data set and used to determine data quality. Cut-off values were calculated using a 99% one-sided confidence interval. Data sets that did not fall within the cut-off values for at least one metric were classed as poor quality.²² Our work presented here differs in that

Table 5. % of signal–time courses passed by each quality control method

Quality control method	% of signal–time courses passed
SDNR threshold	60.97%
RMSE threshold	66.18%
PSR threshold	99.12%
FWHM threshold	97.44%
Random forest ML classifier	75.55%

FWHM, full width half maximum; ML, machine learning; PSR, percentage signal recovery; RMSE, root mean square error; SDNR, signal drop-to-noise ratio.

thresholds for quality are determined using QR results instead of confidence intervals. Combining measures using ML classifiers, leads to improved classification error compared to individual thresholds as shown in Table 4. The random forest classifier gave the lowest classification error, but offers only a minor improvement in performance measures compared to the other classifiers. Therefore, any of the classifiers tested would be suitable.

The ML classifier offers improved sensitivity, classification error and AUC, compared to the SDNR threshold. There is little change in specificity and precision, suggesting that the main improvement in performance comes from a reduction in the number of false negatives, with little change in the number of false positives. ML also has a similar classification rate to the percentage disagreements in QR between reviewers, which implies that it is as accurate as QR at least for these data sets. Therefore, when the quality control methods are applied to a patient data set, the ML classifier passes a higher percentage of signal–time courses than the SDNR threshold, as shown in Figure 8 and Table 5. The lack of reduction in the number of false positives is likely due to the current quantitative measures not being able to identify all the artefacts that DSC-MRI is susceptible to.

The ML classifiers were trained using k-fold validation. Stratified k-fold validation was used to ensure that there was an even distribution of accepted and rejected signal–time courses in each

fold. The centre and data set the signal–time courses originated from was not considered so this was applied on a signal basis rather than a subject basis. This is because the aim of this work is to train a classifier which can be applied to a wide range of patient data, so it needs to be capable of handling data from different centres acquired with different acquisition parameters. Therefore, splitting the data in a subject basis could bias the classifier and reduce its performance.

Currently, the results of qualitative review are applied to a series of quantitative measures which are calculated from the signal–time courses. If a convolutional neural network (CNN) was used in place of the ML classifier, then it could be trained using the signal–time courses directly, rather than calculating measures from the signal–time courses. However, currently there are not enough data to train a CNN type model. This is something that could be investigated in the future once more data have been acquired.

The use of a pre-bolus or single-bolus injection protocol in paediatric data affects the SDNR of the signal–time courses by reducing the signal drop. In adults, a pre-bolus of contrast agent may be given in addition to a full dose of contrast agent, increasing the overall SNR. However, in paediatrics, the European Society for Paediatric Oncology (SIOPE) recommends that paediatric patients should only receive a single-dose of contrast agent, due to concerns over gadolinium deposition.⁴⁴ Splitting a single-dose in order to give a pre-bolus will therefore cause a reduction in SDNR in the DSC-MRI acquisition.

Most DSC-MRI studies use the ASFNr recommendation of QR to assess data quality.¹¹ Automated QC using statistical thresholds has previously been presented by Akella et al.²² Our method differs as the thresholds and ML classifier are trained on the results of QR. An alternative way to assess data quality is for a radiologist to assess the quality of the perfusion maps produced. This could either mean assessing the diagnostic quality of the perfusion maps,^{45,46} assessing the presence of susceptibility artefacts,⁴⁷ or assessing the visibility of a certain region of the brain.⁴⁸ However, these methods are not automated and risk artefacts being misinterpreted as pathology.

In order to calculate the metrics presented in this study, it is necessary to establish the end of the baseline signal in each signal–time course. There are established methods for determining the end of the baseline, *e.g.* Carroll et al.⁴⁹ present a method which uses adaptive thresholds calculated from the standard deviation of the pre-contrast signal, defining a set number of time points from which to calculate the adaptive thresholds.⁴⁹ The method we have presented is better suited to a multicentre data set with variable injection protocols, where the number of dynamics in the baseline may vary between centres.

There are some limitations to this study. Firstly, the patient “training data set does not include every type of artefact, such as susceptibility artefacts or insufficient dynamics to capture the full passage of the contrast agent. There may be cases where the classifier misclassifies a signal–time course with an artefact appearing for

the first time. The training data set is also made up of signal–time courses acquired with specific and consistent acquisition protocols. Any changes in acquisition protocol will affect the signal values and therefore the quantitative measures, *e.g.* PSR can vary with acquisition protocol.¹³ So, thresholds would need to be recalculated at different centres.

Secondly, whilst these methods were tested on signal–time courses from slices of brain that did not contain tumour or other definite pathology, they may still not be “normal tissue”. The quantitative measures may differ in diseased tissue or tissue that has been exposed to treatments such as radiotherapy and so the thresholds and methods presented may not be suitable to all circumstances. For example, PSR has been used to exclude signal–time course with unusual post-bolus signals. However, PSR will be affected by contrast agent leakage due to the breakdown of the blood–brain barrier in tumours. In this case, a leakage correction method, such as the Boxerman-Schmainda-Weisskoff method⁵⁰ could be applied prior to quality assessment. In this study, leakage correction was not applied as no tumour signal–time courses were included.

Finally, although the classifier offers an automated QC method, it is still based on QR so still has an element of subjectivity to it.

CONCLUSIONS

QR of individual signal–time courses by two reviewers showed good agreement on the signal–time courses they assessed. ML classifiers trained on QR results offer an automated method to assess the quality of an entire data set. Although SDNR was a good indicator of quality, using only a single measure to determine data quality risks misclassification of signal–time courses. Combining SDNR with RMSE, FWHM and PSR improves classification, and achieves a misclassification rate similar to the discordance rate of QR. We have shown that ML classifiers trained on QR can be used to assess quality of DSC-MRI signal–time courses obtained from normal brain in this paediatric data set.

ACKNOWLEDGMENT

This work was funded by EPSRC through a studentship from the Sci-Phy-4-Health CDT (EP/L016346/1) and the National Institute for Health Research (NIHR) via a research professorship (RP-R2-12-019). Also, the work has been partially funded by the Children’s Research Fund, Help Harry Help Others (HHHO), Cancer Research UK, NIHR, the Experimental Cancer Medicine Centre Pediatric Network (C8232/A25261), the Children’s Cancer Fund, the Little Princess Trust and HDR UK (HDR-3001). 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 that there are no conflicts of interest.

REFERENCES

- Shiroishi MS, Castellazzi G, Boxerman JL, D'Amore F, Essig M, Nguyen TB, et al. Principles of T2*-weighted dynamic susceptibility contrast MRI technique in brain tumor imaging. *J Magn Reson Imaging* 2015; **41**: 296–313. <https://doi.org/10.1002/jmri.24648>
- Calamante F, Gadian DG, Connelly A. Quantification of perfusion using bolus tracking magnetic resonance imaging in stroke: assumptions, limitations, and potential implications for clinical use. *Stroke* 2002; **33**: 1146–51. <https://doi.org/10.1161/01.str.0000014208.05597.33>
- Calamante F. Perfusion MRI using dynamic-susceptibility contrast MRI: quantification issues in patient studies. *Top Magn Reson Imaging* 2010; **21**: 75–85. <https://doi.org/10.1097/RMR.0b013e31821e53f5>
- van Osch MJP, Barker P, Golay X, Zaharchuk G. n.d. Dynamic susceptibility contrast MRI: acquisition and analysis techniques. *Clinical Perfusion MRI: Techniques and Applications*; **2013**: 16–37.
- Fussell D, Young RJ. Role of MRI perfusion in improving the treatment of brain tumors. *Imaging in Medicine* 2013; **5**: 407–26. <https://doi.org/10.2217/iim.13.50>
- Novak J, Withey SB, Lateef S, MacPherson L, Pinkey B, Peet AC. A comparison of pseudo-continuous arterial spin labelling and dynamic susceptibility contrast MRI with and without contrast agent leakage correction in paediatric brain tumours. *Br J Radiol* 2019; **92**: 20170872. <https://doi.org/10.1259/bjr.20170872>
- Kaufmann TJ, Smits M, Boxerman J, Huang R, Barboriak DP, Weller M, et al. Consensus recommendations for a standardized brain tumor imaging protocol for clinical trials in brain metastases. *Neuro Oncol* 2020; **22**: 757–72. <https://doi.org/10.1093/neuonc/noaa030>
- Hipp SJ, Steffen-Smith E, Hammoud D, Shih JH, Bent R, Warren KE. Predicting outcome of children with diffuse intrinsic pontine gliomas using multiparametric imaging. *Neuro Oncol* 2011; **13**: 904–9. <https://doi.org/10.1093/neuonc/nor076>
- Yeom KW, Mitchell LA, Lober RM, Barnes PD, Vogel H, Fisher PG, et al. Arterial spin-labeled perfusion of pediatric brain tumors. *AJNR Am J Neuroradiol* 2014; **35**: 395–401. <https://doi.org/10.3174/ajnr.A3670>
- Young GS, Stauss J, Mukundan S. Advanced imaging of adult brain tumors with MRI and PET. In: *Blue Books of Neurology*, Vol. 36. Elsevier; 2010, pp. 71–98.
- Welker K, Boxerman J, Kalnin A, Kaufmann T, Shiroishi M, Wintermark M, et al. ASFN recommendations for clinical performance of Mr dynamic susceptibility contrast perfusion imaging of the brain. *AJNR Am J Neuroradiol* 2015; **36**: E41–51. <https://doi.org/10.3174/ajnr.A4341>
- Paulson ES, Prah DE, Schmainda KM. Spiral perfusion imaging with consecutive echoes (SPICESpiral perfusion imaging with consecutive echoes (spiceTMSpiral perfusion imaging with consecutive echoes (spice. *Tomography* 2016; **2**: 295–307. <https://doi.org/10.18383/j.tom.2016.00217>
- Boxerman JL, Paulson ES, Prah MA, Schmainda KM. The effect of pulse sequence parameters and contrast agent dose on percentage signal recovery in DSC-MRI: implications for clinical applications. *AJNR Am J Neuroradiol* 2013; **34**: 1364–69. <https://doi.org/10.3174/ajnr.A3477>
- Boxerman JL, Rosen BR, Weisskoff RM. Signal-To-Noise analysis of cerebral blood volume maps from dynamic NMR imaging studies. *J Magn Reson Imaging* 1997; **7**: 528–37. <https://doi.org/10.1002/jmri.1880070313>
- Zhang J, Liu H, Tong H, Wang S, Yang Y, Liu G, et al. Clinical applications of contrast-enhanced perfusion MRI techniques in gliomas: recent advances and current challenges. *Contrast Media Mol Imaging* 2017; **2017**: 7064120. <https://doi.org/10.1155/2017/7064120>
- Willats L, Calamante F. The 39 steps: evading error and deciphering the secrets for accurate dynamic susceptibility contrast MRI. *NMR Biomed* 2013; **26**: 913–31. <https://doi.org/10.1002/nbm.2833>
- Boxerman JL, Prah DE, Paulson ES, Machan JT, Bedekar D, Schmainda KM. The role of preload and leakage correction in gadolinium-based cerebral blood volume estimation determined by comparison with MION as a criterion standard. *Am J Neuroradiol* 2012; **33**: 1081–87. <https://doi.org/10.3174/ajnr.A2934>
- Hu LS, Baxter LC, Pinnaduwa DS, Paine TL, Karis JP, Feuerstein BG, et al. Optimized preload leakage-correction methods to improve the diagnostic accuracy of dynamic susceptibility-weighted contrast-enhanced perfusion MR imaging in posttreatment gliomas. *Am J Neuroradiol* 2010; **31**: 40–48. <https://doi.org/10.3174/ajnr.A1787>
- Boxerman JL, Quarles CC, Hu LS, Erickson BJ, Gerstner ER, Smits M, et al. Consensus recommendations for a dynamic susceptibility contrast MRI protocol for use in high-grade gliomas. *Neuro Oncol* 2020; **22**: 1262–75. <https://doi.org/10.1093/neuonc/noaa141>
- Schmainda KM, Prah MA, Hu LS, Quarles CC, Semmineh N, Rand SD, et al. Moving toward a consensus DSC-MRI protocol: validation of a low-flip angle single-dose option as a reference standard for brain tumors. *Am J Neuroradiol* 2019; **40**: 626–33. <https://doi.org/10.3174/ajnr.A6015>
- Christoforidis GA. Ultra-High-Field Magnetic Resonance Imaging of Brain Tumors. In: *Handbook of Neuro-Oncology Neuroimaging*. Elsevier; 2016, pp. 385–94.
- Akella NS, Twieg DB, Mikkelsen T, Hochberg FH, Grossman S, Cloud GA, et al. Assessment of brain tumor angiogenesis inhibitors using perfusion magnetic resonance imaging: quality and analysis results of a phase I trial. *J Magn Reson Imaging* 2004; **20**: 913–22. <https://doi.org/10.1002/jmri.20202>
- Al-Hashem MA, Alqudah AM, Qananwah Q. Performance evaluation of different machine learning classification algorithms for disease diagnosis. *International Journal of E-Health and Medical Communications* 2021; **12**: 1–28. <https://doi.org/10.4018/IJEHMC.20211101.0a5>
- S. Masad I, Alqudah A, Alqudah AM, Almashaqbeh S. A hybrid deep learning approach towards building an intelligent system for pneumonia detection in chest X-ray images. *IJECE* 2088; **11**: 5530. <https://doi.org/10.11591/ijece.v11i6.pp5530-5540>
- Alqudah AM, Qazan S, Masad IS. Artificial intelligence framework for efficient detection and classification of pneumonia using chest radiography images. *J Med Biol Eng* 2021; **41**: 599–609. <https://doi.org/10.1007/s40846-021-00631-1>
- Alquran H, Alsletti M, Alsharif R, Abu Qasmieh I, Alqudah AM, Binti Harun NH. Employing texture features of chest X-ray images and machine learning in covid-19 detection and classification. *Mendel* 2021; **27**: 9–17. <https://doi.org/10.13164/mendel.2021.1.009>
- Alqudah AM, Alqudah A. Improving machine learning recognition of colorectal cancer using 3D GLCM applied to different color spaces. *Multimed Tools Appl* 2022; **81**: 10839–60. <https://doi.org/10.1007/s11042-022-11946-9>
- Alqudah AM, Al-Hashem M, Alqudah A. Reduced number of parameters for predicting post-stroke activities of daily living using machine learning algorithms on

- initiating rehabilitation. *IJCAI* 2021; **45**(4). <https://doi.org/10.31449/inf.v45i4.3570>
29. McKinley R, Hung F, Wiest R, Liebeskind DS, Scalzo F. A machine learning approach to perfusion imaging with dynamic susceptibility contrast mr. *Front Neurol* 2018; **9**: 717. <https://doi.org/10.3389/fneur.2018.00717>
 30. Emblem KE, Due-Tonnessen P, Hald JK, Bjornerud A, Pinho MC, Scheie D, et al. Machine learning in preoperative glioma MRI: survival associations by perfusion-based support vector machine outperforms traditional MRI. *J Magn Reson Imaging* 2014; **40**: 47–54. <https://doi.org/10.1002/jmri.24390>
 31. Jeong J, Wang L, Ji B, Lei Y, Ali A, Liu T, et al. Machine-learning based classification of glioblastoma using delta-radiomic features derived from dynamic susceptibility contrast enhanced magnetic resonance images: introduction. *Quant Imaging Med Surg* 2019; **9**: 1201–13. <https://doi.org/10.21037/qims.2019.07.01>
 32. Ji B, Wang S, Liu Z, Weinberg BD, Yang X, Liu T, et al. Revealing hemodynamic heterogeneity of gliomas based on signal profile features of dynamic susceptibility contrast-enhanced MRI. *NeuroImage: Clinical* 2019; **23**: 101864. <https://doi.org/10.1016/j.nicl.2019.101864>
 33. Grist JT, Withey S, MacPherson L, Oates A, Powell S, Novak J, et al. Distinguishing between paediatric brain tumour types using multi-parametric magnetic resonance imaging and machine learning: a multi-site study. *Neuroimage Clin* 2020; **25**: 102172. <https://doi.org/10.1016/j.nicl.2020.102172>
 34. Hu LS, Yoon H, Eschbacher JM, Baxter LC, Dueck AC, Nespodzany A, et al. Accurate patient-specific machine learning models of glioblastoma invasion using transfer learning. *Am J Neuroradiol* 2019; **40**: 418–25. <https://doi.org/10.3174/ajnr.A5981>
 35. Tsolaki E, Svolos P, Kousi E, Kapsalaki E, Fountas K, Theodorou K, et al. Automated differentiation of glioblastomas from intracranial metastases using 3T Mr spectroscopic and perfusion data. *Int J Comput Assist Radiol Surg* 2013; **8**: 751–61. <https://doi.org/10.1007/s11548-012-0808-0>
 36. Swinburne NC, Schefflein J, Sakai Y, Oermann EK, Titano JJ, Chen I, et al. Machine learning for semi-automated classification of glioblastoma, brain metastasis and central nervous system lymphoma using magnetic resonance advanced imaging. *Ann Transl Med* 2019; **7**: 232. <https://doi.org/10.21037/atm.2018.08.05>
 37. Arvanitis T, Natarajan K, Rossiter J, Ting J, Sun Y, Wilson M, et al. The Children's Cancer and Leukaemia Group (CCLG) functional imaging e-repository for clinical trials of childhood brain tumours. *Neuro Oncol*. 2010;12(6).
 38. Pedersen M, Klarhöfer M, Christensen S, Ouallet J-C, Ostergaard L, Dousset V, et al. Quantitative cerebral perfusion using the PRESTO acquisition scheme. *J Magn Reson Imaging* 2004; **20**: 930–40. <https://doi.org/10.1002/jmri.20206>
 39. Digernes I, Nilsen LB, Grøvik E, Bjørnerud A, Løvland G, Vik-Mo E, et al. Noise dependency in vascular parameters from combined gradient-echo and spin-echo DSC MRI. *Phys Med Biol* 2020; **65**: 225020. <https://doi.org/10.1088/1361-6560/abb71a>
 40. Madsen MT. A simplified formulation of the gamma variate function. *Phys Med Biol* 1992; **37**: 1597–1600. <https://doi.org/10.1088/0031-9155/37/7/010>
 41. Jung Y. Multiple predicting K -fold cross-validation for model selection . *Journal of Nonparametric Statistics* 2018; **30**: 197–215. <https://doi.org/10.1080/10485252.2017.1404598>
 42. Paluszek M, Thomas S. *MATLAB Machine Learning*. Apress; 2016. <https://doi.org/10.1007/978-1-4842-2250-8>
 43. McHugh ML. Interrater reliability: the kappa statistic. *Biochem Med* 2012; **22**: 276–82. <https://doi.org/10.11613/BM.2012.031>
 44. Avula S, Peet A, Morana G, Morgan P, Warmuth-Metz M, Jaspan T, et al. European Society for paediatric oncology (SIOPE) MRI guidelines for imaging patients with central nervous system tumours. *Childs Nerv Syst* 2021; **37**: 2497–2508. <https://doi.org/10.1007/s00381-021-05199-4>
 45. Armitage PA, Skipper N, Connolly DJA, Griffiths PD. A qualitative comparison of arterial spin labelling and dynamic susceptibility contrast MRI in 52 children with A range of neurological conditions. *Br J Radiol* 2017; **90**: 20160495. <https://doi.org/10.1259/bjr.20160495>
 46. Thilmann O, Larsson E-M, Björkman-Burtscher IM, Ståhlberg F, Wirestam R. Effects of echo time variation on perfusion assessment using dynamic susceptibility contrast MR imaging at 3 tesla. *Magn Reson Imaging* 2004; **22**: 929–35. <https://doi.org/10.1016/j.mri.2004.01.079>
 47. Xu Q, Liu Q, Ge H, Ge X, Wu J, Qu J, et al. Tumor recurrence versus treatment effects in glioma: a comparative study of three dimensional pseudo-continuous arterial spin labeling and dynamic susceptibility contrast imaging. *Medicine (Baltimore)* 2017; **96**: e9332. <https://doi.org/10.1097/MD.00000000000009332>
 48. Varallyay CG, Nesbit E, Horvath A, Varallyay P, Fu R, Gahramanov S, et al. Cerebral blood volume mapping with ferumoxytol in dynamic susceptibility contrast perfusion MRI: comparison to standard of care. *J Magn Reson Imaging* 2018; **48**: 441–48. <https://doi.org/10.1002/jmri.25943>
 49. Carroll TJ, Rowley HA, Haughton VM. Automatic calculation of the arterial input function for cerebral perfusion imaging with MR imaging. *Radiology* 2003; **227**: 593–600. <https://doi.org/10.1148/radiol.2272020092>
 50. Boxerman JL, Schmainda KM, Weisskoff RM. Relative cerebral blood volume maps corrected for contrast agent extravasation significantly correlate with glioma tumor grade, whereas uncorrected maps do not. *Am J Neuroradiol* 2006; **27**: 859–67.