Identifying quantitative imaging features of Posterior Fossa Syndrome in longitudinal MRI

Michaela Spiteri[a], David Windridge[a], Shivaram Avula[b], Ram Kumar[b], Emma Lewis[a]
[a]Centre for Vision, Speech and Signal Processing, University of Surrey, Guildford, UK, GU2 7XH
[b]Alder Hey Children's Hospital (NHS), Eaton Rd, Liverpool, UK, L12 2AP

Abstract. Up to 25% of children who undergo brain tumour resection surgery in the posterior fossa develop posterior fossa syndrome (PFS). This syndrome is characterised by mutism and disturbance in speech. Our hypothesis is that there is a correlation between PFS and the occurrence of hypertrophic olivary degeneration (HOD) in structures within the posterior fossa, known as the inferior olivary nuclei (ION). HOD is exhibited as an increase in size and intensity of the ION on an MR image.

Longitudinal MRI datasets of 28 patients were acquired consisting of pre, intra and post operative scans. A semi-automated segmentation process was used to segment the ION on each MR image. A full set of imaging features describing the first and second order statistics and size of the ION were extracted for each image. Feature selection techniques were used to identify the most relevant features amongst the MR imaging features,demographics and data based on neuroradiological assessment. A support vector machine (SVM) was used to analyse the discriminative features selected by a generative nearest neighbour algorithm (k-NN). The results indicate the presence of hyperintensity in the left ION as the most diagnostically relevant feature, providing a statistically significant improvement in the classification of patients (p=0.01) when using this feature alone.

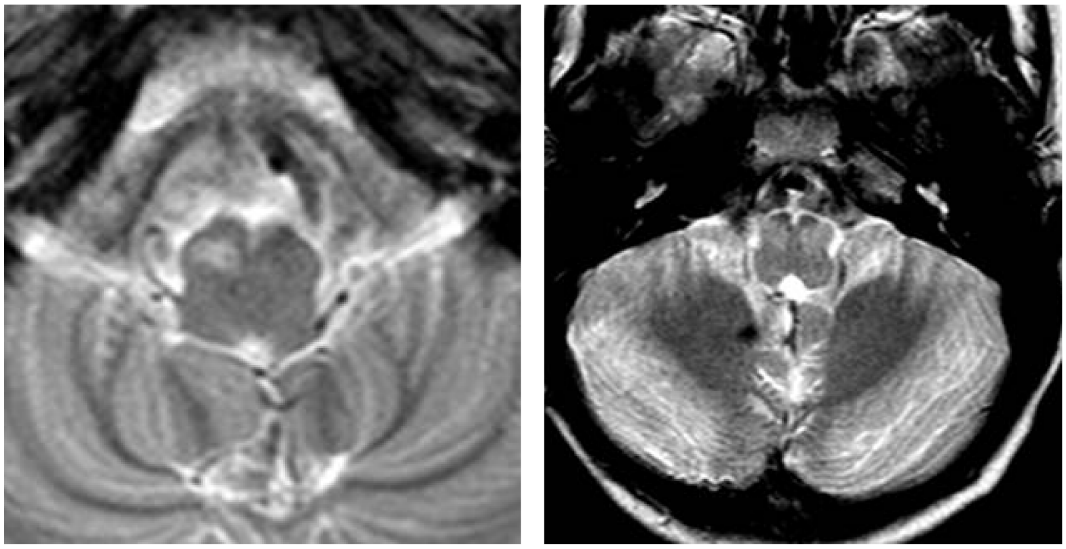
Keywords: Posterior Fossa Syndrome, Inferior Olivary Nuclei, Hypertrophic Olivary Degeneration, Intra-operative MRI.

Address all correspondence to: Michaela Spiteri, Centre for Vision, Speech and Signal Processing, University of Surrey, Guildford, UK, GU2 7XH; Tel: $+44(0)1483689856$; E-mail: m.spiteri@surrey.ac.uk

1 Introduction

The posterior fossa is the commonest site for intracranial tumours in children. Up to 1 in 4 children who undergo brain tumour resection surgery in the posterior fossa develop a syndrome known as Posterior Fossa Syndrome (PFS).[1] This syndrome, also known as cerebellar mutism syndrome (CMS), describes a set of neurological symptoms which may develop from 24 to 107 hours after surgery.[2,3] Children suffering from PFS, characteristically suffer from disturbance in speech and mutism, but may also suffer from loss of muscle tone, incontinence, strabismus (cross-eyed), dysphagia, and personality changes such as anger, apathy, melancholy, crying and screaming.[2] The development of such a syndrome in children hinders their development and highly impacts their quality of life. Although PFS is a post surgical complication, the exact underlying pathophysiological mechanism remains unclear, although it is widely considered to involve disruption of the proximal efferent cerebellar pathways (pECP) that connect the cerebellum to the forebrain. In order to reduce the incidence of PFS and manage children with this disorder, it is important to identify imaging biomarkers that are associated with it.

Our hypothesis, based on qualitative interpretation of imaging and clinical experience, is that there is a correlation between PFS and the occurrence of hypertrophic olivary degeneration (HOD) in structures known as the inferior olivary nuclei (ION). These structures, shown in Figure 1, are paired nuclei in the brain stem which send efferent outputs to the cerebellum, and receive inputs from the pECP. HOD is exhibited as an increase in size and intensity of the ION on an MR image



(a) (b)

Fig 1: Hypertrophic olivary degeneration: (a) unilateral case and (b) bilateral case.

which in routine clinical practice is identified qualitatively by a neuroradiologist.[4] Qualitatively HOD appears a number of months after surgery on routine post-operative surveillance imaging, that is well after the patient is diagnosed with PFS. We thus hypothesise that HOD represents the "smoking gun" that a preceding damaging event to the relevant pathways connected to the ION has occurred.

Intra-operative MRI is increasingly used to ensure the tumour is removed safely, for example in the resection of posterior fossa tumours. Intra-operative MRI (IoMRI) is used during posterior fossa tumour resection. The use of IoMRI increases the likelihood that the entire tumour is removed and hence increases the likelihood of success after surgery[5,6] but PFS can occur after attempting total resection of tumour causing injury to important structures. Furthermore, the final MR scan acquired using IoMRI provides quantitative information about the state of the ION immediately after the surgical procedure.

In the present study, we propose the quantification of HOD using longitudinal imaging features in the aim of identifying imaging features that correlate with the incidence of PFS in children. The aim was to analyse and compare imaging features in the ION on a longitudinal MRI dataset with the intention of establishing a link between PFS and HOD. Association of HOD and PFS will add to the existing evidence on the development of PFS and potentially lead to a deeper understanding of the pathogenesis of the syndrome. Segmentation was applied to the IONs on each image in the longitudinal datasets and quantitative features were chosen to describe longitudinal changes in the area and intensity of the left and right ION. Feature selection techniques were applied to these features in order to identify the optimal feature set. A classification model was applied to the original feature set as well as the optimised feature subsets to demonstrate the improvement in classification accuracy when using the optimised feature subsets.

Table 1: Patient longitudinal dataset

Pt	Image acquisition: days after surgery						Gender	Age	HOD	Bi-l	Uni-l	PFS
1	1	110	194	/	/	/	M	4	0	0	0	0
2	121	205	289	401	/	/	F	7	1	1	0	0
3	120	204	288	400	/	/	M	6	0	0	0	0
4	1	77	118	/	/	/	F	0	0	0	0	0
5	1	98	231	413	/	/	F	4	0	0	0	0
6	-30	0	175	287	403	/	F	4	0	0	0	0
7	143	318	437	/	/	/	M	3	1	1	0	0
8	118	278	481	/	/	/	M	7	1	0	1	0
9	176	260	372	/	/	/	F	7	1	1	0	1
10	189	273	357	/	/	/	F	7	1	0	1	0
11	-1	2	96	193	216	334	M	3	0	0	0	0
12	91	228	351	/	/	/	F	8	1	1	0	1
13	18	21	165	228	/	/	F	14	1	0	1	0
14	IO_1	76	87	28	60	/	F	2	0	0	0	0
15	136	257	440	/	/	/	F	5	1	1	0	0
16	89	285	/	/	/	/	M	1	0	0	0	0
17	313	481	/	/	/	/	F	6	1	1	0	1
18	IO_1	110	446	/	/	/	F	10	0	0	0	0
19	124	288	481	/	/	/	M	17	1	1	0	1
20	97	181	321	/	/	/	F	12	1	1	0	1
21	IO_1	IO_2	108	255	445	/	M	11	0	0	0	0
22	IO_1	IO_2	125	/	/	/	M	3	0	0	0	0
23	131	299	/	/	/	/	F	15	1	1	0	1
24	IO_1	19	201	322	/	/	F	14	0	0	0	0
25	-2	IO_1	IO_2	173	509	/	M	13	0	0	0	0
26	184	228	/	/	/	/	F	3	1	0	1	1
27	IO_1	IO_2	178	273	424	/	M	8	0	0	0	1
28	IO_1	1	110	292	/	/	F	7	0	0	0	1

Key		
HOD	Presence of HOD	1 = yes, 0 = no
Bi-l	Presence of Bilateral HOD	1 = yes, 0 = no
Uni-l	Presence of Unilateral HOD	1 = yes, 0 = no
PFS	PFS diagnosis	1 = yes, 0 = no

2 Study Dataset

The dataset was compiled from 28 of patients treated for various histological types of posterior fossa tumours at Alder Hey Children's Hospital between 2007 and 2013. The patients were aged between 8 months and 18 years old (at surgery), nine of whom were diagnosed with PFS as reported qualitatively by a consultant neuroradiologist who was blinded to the child's neurological condition. There exist two schema for diagnosis: sensitive PFS and specific PFS. This study is

based on correlating HOD to the specific diagnosis, however it is worth noting that two additional patients within the dataset were also diagnosed with sensitive PFS. The methodology of the study did not fulfil the criteria for research ethics approval and was given institutional approval by the Director of Research at Alder Hey Children's Hospital.

Thirteen of these patients exhibited HOD, nine bilaterally (in both ION) and four unilaterally (in either the left or right ION). Follow up MR images during one year post-surgery were reviewed and up to five MR images were acquired longitudinally for each patient across their treatment. A small subset of these datasets included intra-operative MR images. Table 1 describes the MR dataset acquired for each patient, showing the number of days after surgery when an MR image was acquired. Negative numbers indicate a pre-operative MR image, whilst intra-operative scans are indicated as IO_1 or IO_2. Intra-operative scans were acquired to determine whether the surgical aim had been achieved. If the surgical aim was not achieved some patients had further resections and intra-operative scans. The first intra-operative scan is indicated as IO_1 and the second intra-operative scan is indicated as IO_2. Pre-operative scans acquired on the day of the surgery are indicated as 0. Most patients were followed-up every three months, whilst others with potentially malignant tumours were followed-up more frequently. A mean of 4 ± 1 MR images were acquired for each patient with a mean time interval of 109 ± 62 days between each image acquisition. The age column refers to the patient's age at surgery. Patients who were diagnosed with HOD (as determined by expert radiological assessment) are indicated as a 1, whilst those who did not develop HOD are indicated as a 0. Similarly for whether HOD occurred bilaterally (Bi-l) or unilaterally (Uni-l), and whether the patient was clinically diagnosed with PFS by a neurologist.

T2 weighted sequences from the pre-, intra- and post-operative scan were used to evaluate for HOD and the following parameters were used: TR = 4485 ms, TE = 11ms, slice thickness = 6mm, number of slices = 20, time-step = 4.49 s. The pre-operative and post-operative MR images were acquired using 1.5T or 3T magnets. Intra-operative MR images were acquired using 3T magnets. This modality was used due to its ability to identify cerebrospinal fluid, blood and edema as increased grey-level intensity. The T1 MR images obtained for these patients were not analysed as they do not provide sufficient information relating to hypertrophy in the ION.

T2 volumetric imaging is not routinely used as it is time consuming and prone to movement associated artefacts. Instead, axial T2 spin-echo sequences were used to evaluate for HOD as they result in the best signal and contrast resolution to assess the ION. These T2 MR images were acquired in Spiral MRI which captures the k-space through a spiral trajectory. This method of acquisition is fast and results in high in-plane spatial resolution, giving improved resolution of small structures within the brain, specifically the ION.[7,8]

3 Methodology

The aim of this study is to identify bio-markers that correlate with the development of PFS following tumour resection surgery in the posterior fossa. In order that these bio-markers may aid understanding of the pathogenesis of PFS, techniques have been chosen to ensure that comprehensibility of imaging and clinical features is retained throughout the pipeline. This study consists of four stages: Image Pre-Processing, Feature Extraction, Feature Selection and Classification. The features were chosen to quantify HOD, namely an increase in intensity and size, in the left or right ION.

3.1 Image Pre-Processing

In order to extract information (features) about each ION it was necessary to segment these structures on the MR images. Images were acquired with spiral MRI, therefore a full volumetric representation was not obtained. For this reason segmentation was performed on two-dimensional image slices.

The non-HOD ION cannot be clearly delineated by the naked eye on MRI. This is due to very low contrast with the surrounding tissue as well as its relatively small cross-sectional area. For this reason, segmentation was carried out using a semi-automated seed-growing technique in two-dimensional space. The right and left ION were segmented separately. Images were registered to Talaraich space, using a rigid body affine transformation, with intensity scaling, prior to segmentation.

The process by which segmentation is carried out consists of three main steps: (1) an arbitrary seed-point within the ION was manually identified using prior anatomical knowledge - the IONs are on the anterior part of the brain stem, located on either side of its mid-line (when the IONs are hypertrophic their grey-level intensity is relatively higher than surrounding brain stem tissue and are therefore easier to identify) (2) region growing from a seed-point, with intensity I_s, was performed by analysing pixels in a search space of a 4mm radius: a pixel within the search space is included in the region of interest (ROI) if its grey-level intensity, I_p, lies within the range $|I_p \pm T| \leq |I_s|$, and its difference from adjacent pixels, I_a, lies within the range $|I_a| \leq |I_p \pm T|$, where T is a threshold that was varied between 12 and 16 heuristically until the ROI did not vary in shape or size;[9] (3) the application of a morphological closing operation using a full width at half maximum of 4mm and a threshold of 0.5.[9]

These steps are applied iteratively until no further change occurs in the region of interest. The segmentation process was carried out three times for each MR image in order to assess intra-observer variability. The first segmentation dataset was validated and amended by an expert neuroradiologist. This introduced a measure of inter-observer variability as the first segmentation test set was expertly validated, whilst the other two segmentation test sets were not.

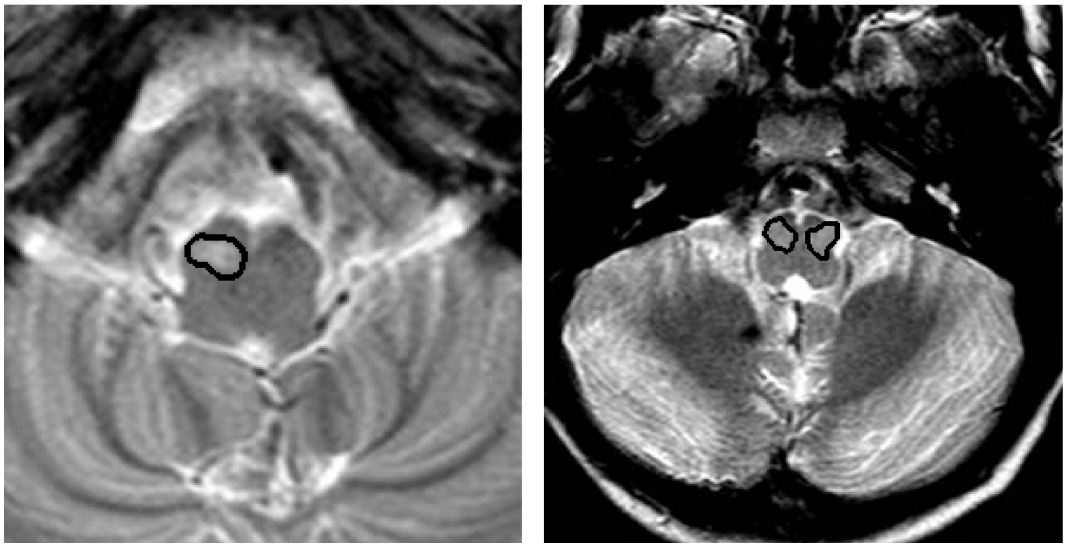
3.2 Feature Extraction

Once the desired region was segmented it was possible to extract a set of features from each ION. HOD is characterised by an increase in volume of the ION which can be seen as both an enlargement and an increase in signal intensity on a T2-weighted MR image. Imaging features related to an increase in size and image intensity are extracted from the MR images. The area of the left and right IONs are obtained as well as the contrast between the left and right ION and surrounding brain stem tissue within the same MR image slice.

The contrast was calculated using the definition of Weber contrast (Equation 1) where I_{ION}refers to the mean grey level intensity of the ION and I_b refers to the mean grey level intensity of the surrounding tissue.

$$W = \frac{I_{ION} - I_b}{I_b} \tag{1}$$

For each MRI, the contrast of both the left and right ION was calculated separately. The segmentation of the left and right ION is exhibited in Figure 2.



(a) (b)

Fig 2: Segmentation of Inferior Olivary Nuclei (delineated in black): (a) unilateral case and (b) bilateral case.

The imaging features are chosen to relate to physiological characteristics of HOD, namely a change in intensity with respect to surrounding brain stem tissue and an increase in area of the left and right ION respectively. It was desired to quantify these characteristics longitudinally.

The contrast, defined in Eq 1, for both the left and right ION, C_L and C_R, was obtained for up to 6 MR images per patient acquired at different time points throughout each patient's treatment. The mean gradient of contrast against time was calculated, symbolised by $mean(\frac{\Delta C_L}{\Delta t})$ and $mean(\frac{\Delta C_R}{\Delta t})$, respectively. The variance of gradient of contrast against time, $var(\frac{\Delta C_L}{\Delta t})$ and $var(\frac{\Delta C_R}{\Delta t})$, was also calculated across each patient's longitudinal set of MR Images.

Similarly the area of the ION was calculated from each MRI and the mean slope and variance across each longitudinal dataset for the left and right ION separately. These values are symbolised by $mean(\frac{\Delta A_L}{\Delta t})$, $mean(\frac{\Delta A_R}{\Delta t})$, $var(\frac{\Delta A_L}{\Delta t})$ and $var(\frac{\Delta A_R}{\Delta t})$.

Features determined by expert radiological assessment of each MR image were also included, namely whether HOD is present (1) or not (0), whether HOD is present unilaterally (1) or not (0) and whether HOD is present bilaterally (1) or not (0). The neuroradiologist was blinded to the PFS status of the patient. It is important to note that these features are not mutually exclusive and the lack of presence of HOD bilaterally may imply either unilateral HOD or no HOD.

The features included are shown in table 2. Features (1) to (8) are obtained from MR data, features (9) to (14) represent clinical data. Features (15) and (16) represent random noise added in order to assess the discriminative ability of the feature selection algorithms used later.

3.3 Feature Selection

Dimensionality reduction techniques, such as Principal Component Analysis (PCA), result in loss of comprehensibility from the point of view of a clinical practitioner,[10] rendering it inappropriate for this application due to the need for medics and clinicians to interpret results. PCA identifies

Table 2: Features included

Feature	Definition
1	$mean(\Delta C_L/\Delta t)$
2	$var(\Delta C_L/\Delta t)$
3	$mean(\Delta A_L/\Delta t)$
4	$var(\Delta A_L/\Delta t)$
5	$mean(\Delta C_R/\Delta t)$
6	$var(\Delta C_R/\Delta t)$
7	$mean(\Delta A_R/\Delta t)$
8	$var(\Delta A_R/\Delta t)$

Feature	Definition
9	Presence of HOD
10	Bilateral HOD
11	Unilateral HOD
12	Enlargement
13	Gender
14	Age at Surgery
15	Random Noise 1
16	Random Noise 2

linear combinations of features as opposed to discrete ones and is therefore less applicable to diagnosis.

From a machine learning respective, to avoid the classifier over fitting the data, in the case of too many features, it is desirable to use only the most relevant features in classifying data into two groups: patients who have developed PFS and those who have not. We however, have an additional motivation; the determination of diagnostically-relevant medical indicators. This is known as Feature Selection and can be carried out using Filter or Wrapper Methods.[11,12]

In general, the problem of feature selection is NP-hard, and therefore intractable for large datasets. Various techniques have therefore been applied, however, these are all prone to local minima. The most common techniques used to identify the salient features out of the full feature set are: random subset feature selection (RSFS), sequential forward selection (SFS) and sequential floating forward selection (SFFS).[13,14]

For each feature selection algorithm a subset of features is chosen and classification is carried out as a criterion for selecting the optimal features. A k-NN classifier was used in each algorithm as it is a generative technique which follows the underlying distribution of data. A support vector machine (used for classification in Section 3.4) was not ideal for this task as it is a discriminative technique and hence more ideal for binary diagnostic classification. The relevance of each feature was scored using a UAR as in the case of the RSFS algorithm.[13,14]

RSFS chooses a random subset of features from the entire feature set, the size of which is equal to the square root of the total number of features. A k-NN classification using three neighbours is carried out repeatedly on this chosen subset. Each feature is given a relevance score which is continuously updated according to its inclusion in the random subsets which perform well.[13,15] The relevance values of each feature are compared to random walk statistics and good features are chosen accordingly. The algorithm is carried out until the stopping criterion is reached, that is, if the size of the final feature set (consisting of the features with the highest relevance scores) has not changed by more than 0.5% in the previous 1000 iterations, or if the maximum number of iterations (300, 000) is reached. The RSFS algorithm was carried out 100 times, each time randomly dividing the dataset in two.

Unlike RSFS, SFS starts off with an empty data set. One feature is added at a time and a feature is kept or discarded depending on whether it exhibits the best classification performance when used together with the previously chosen features. SFS also makes use of k-NN classifier on the feature subset in order to obtain a classification score. Low-scoring features were discarded.

In SFFS an attempt is made at finding the least useful feature in order to discard it from the final feature set. This process is repeated until the evaluation score becomes (and remains) better than the previous best score using a feature set of the same size.[13,14] Both the SFS and SFFS algorithms were carried out using 3 neighbours, 4 neighbours, 5 neighbours and 6 neighbours. This process was carried out 100 times and the average relevance scores were calculated.

All three feature selection methods were carried out on three separate segmentation test sets in order to assess differences in scores that may arise due to intra-observer and inter-observer variability.

3.4 Classification

The binary classification was carried out in order to assess the discriminative ability of the most relevant features chosen in the previous stage of the study. The aim is to classify patients into two groups: patients who developed PFS and patients who had not developed PFS.

Two different feature subsets were used; the first subset included the entire feature set whilst the second subset included the most relevant features chosen by the RSFS, the SFS and SFFS algorithms. A simple linear non-kernelised support vector machine (SVM) was used to perform the classification task. Support Vector Machines (SVMs) is a state of the art classification model used for binary classification when the dataset falls into two main categories (SVMs).[16,17]

Due to the small size of the dataset it was not feasible to split the data into training data and test data. Since there is no natural division between training and test sets within the data, the most efficient and ideal way to maximise the use of this small dataset, was to implement a leave-M-out cross-validation (LMOCV). In this validation technique M observations are omitted from the entire set for training purposes; the M observations are then used as the test set; this process is repeated a number of times in order to obtain a mean value for the area under the curve (AUC) and the accuracy of the SVM classifier. A leave-8-out cross-validation for each feature subset was carried out 100,000 times. For each permutation the SVM bias was varied between -4 and 4, in increments of 0.2. The false positives and the false negatives were obtained for each bias point, and a mean of these values across all the permutations was obtained. Receiver operating characteristic (ROC) graphs, exhibiting the false positives against the true positives, were plotted in order to assess the difference in classification accuracy when using different feature subsets; the area under the ROC curves was obtained. A leave-one-out cross-validation was carried out for all the patients (28 times) in order to obtain a mean accuracy score for each feature subset.

4 Results

4.1 Features

Table 3 exhibits the mean, μ and standard deviation, σ, for features 1 to 8, for segmentation test sets 1,2 and 3.

4.2 Feature Selection

Table 4 shows the relevance scores calculated by the random subset feature selection (RSFS) algorithm. Table 5 displays the relevance scores calculated by the sequential forward selection (SFS) algorithm and Table 6 displays the relevance scores calculated by the sequential floating forward (SFFS) algorithm using a k-NN classifier with 3, 4, 5 and 6 neighbours; these algorithms yielded

Table 3: The mean and standard deviation of the imaging features, for segmentation test sets 1,2 and 3

$\mu \pm \sigma$(x10^{-4})				
Feature	1	2	3	4
Test	$mean(\Delta C_L/\Delta t)$	$var(\Delta C_L/\Delta t)$	$mean(\Delta A_L/\Delta t)$	$var(\Delta A_L/\Delta t)$
1	-2.26 ± 4.17	75.48 ± 82.89	0.51 ± 1.41	10.17 ± 19.58
2	-1.64 ± 3.19	80.13 ± 91.00	-0.23 ± 0.78	3.59 ± 3.36
3	-1.21 ± 4.45	66.30 ± 56.75	0.85 ± 1.18	7.43 ± 11.64

$\mu \pm \sigma$(x10^{-4})				
Feature	5	6	7	8
Test	$mean(\Delta C_R/\Delta t)$	$var(\Delta C_R/\Delta t)$	$mean(\Delta A_R/\Delta t)$	$var(\Delta A_R/\Delta t)$
1	-1.20 ± 5.64	105.49 ± 96.98	0.63 ± 1.49	12.40 ± 25.55
2	1.16 ± 6.35	89.29 ± 107.60	0.04 ± 0.93	3.79 ± 2.85
3	-0.20 ± 5.08	86.10 ± 66.81	0.34 ± 1.14	6.85 ± 12.13

Table 4: The relevance scores calculated by the random subset feature selection algorithm

Average Relevance Score over 100 iterations								
	Feature Key							
Test	1	2	3	4	5	6	7	8
1	6.22	1.61	1.08	1.81	1.64	0.23	0.65	0.11
2	8.99	0.09	0.31	1.41	1.03	0.20	0.28	1.22
3	10.03	1.24	0.00	0.42	0.00	0.12	1.62	0.15

Average Relevance Score over 100 iterations								
	Feature Key							
Test	9	10	11	12	13	14	15	16
1	1.06	0.94	0.16	0.03	0.42	0.09	0.00	0.00
2	0.92	1.77	0.27	0.10	1.28	0.30	0.00	0.00
3	1.19	1.46	0.00	0.00	0.18	0.59	0.00	0.00

identical results. Feature 1 consistently obtained the highest score for all feature selection techniques.

4.3 Classification

Figure 3 exhibits the receiver operating characteristic curve for the SVM classifier used on the full feature dataset and feature 1, the most relevant feature found using the RSFS, SFS and SFFS algorithms.

Table 5: The average relevance scores calculated by the sequential forward selection algorithm over 100 iterations

Average Relevance Score over 100 iterations					
Test	Feature Key	k=3	k=4	k=5	k=6
1	1	68.38	72.68	74.10	71.37
2	1	77.93	76.04	73.37	69.77
3	1	85.10	85.31	85.87	86.34

Table 6: The average relevance scores calculated by the sequential floating forward selection algorithm over 100 iterations

Average Relevance Score over 100 iterations					
Test	Feature Key	k=3	k=4	k=5	k=6
1	1	68.88	75.32	73.88	73.22
2	1	76.75	77.56	74.34	71.48
3	1	84.98	85.18	85.87	86.39

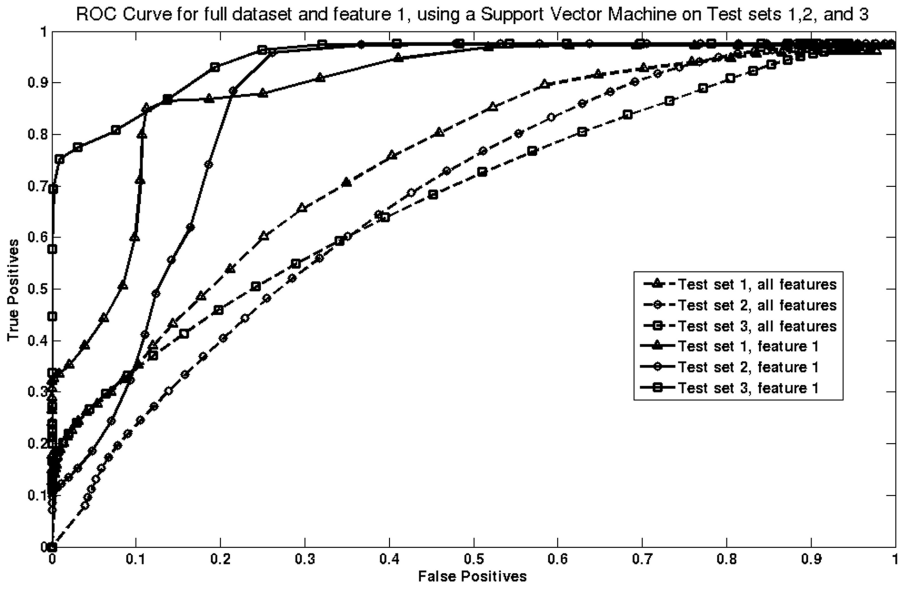


Fig 3: The receiver operating characteristic (ROC) curve for the SVM classifier used on: the full feature dataset and feature 1 on Test sets 1,2, and 3

Table 7 reports the area under the curve (AUC) for the SVM classifier for all three segmentation test sets carried out on the full feature set and feature 1, as well as the average AUC and accuracy for the full feature set and feature 1. A paired t-test was carried out to calculate the significance of the difference between the AUCs of the three ROC curves when considering feature 1, and the three ROC curves when considering the whole feature set for test sets 1,2, and 3. The confidence

Table 7: Area under the curve and accuracy for the SVM classifier used on: the full feature dataset and feature 1 on Test sets 1,2, and 3

	Area Under the Curve		Accuracy (%)	
Test	Full Feature Set	Feature 1	Full Feature Set	Feature 1
1	0.74	0.89	78.57	89.29
2	0.62	0.85	75.00	78.57
3	0.64	0.89	71.43	85.71
Average	0.67	0.88	75.00	84.52

interval was taken to be 95%. The two-tailed p-value was found to be 0.01.

5 Discussion

The results yielded by the RSFS algorithm in Table 4 indicate feature 1 as the most relevant feature, scoring higher than all other features considered in this study. This feature corresponds to the mean slope of contrast in the left nucleus. The score for this feature in each test set was 6.22, 8.99, and 10.03 for test sets 1,2, and 3, respectively. These scores are at least 5 times higher than the scores for all the other features in the full feature set.

These results indicate changes in contrast in the left ION as the most relevant feature correlating with the development of PFS. This implies that change in intensity of the left ION as seen on MRI is highly correlated to the presence of PFS. This quantified contrast in the left ION from patient MR is at least six times as predictive as the diagnosis of HOD made by radiological assessment as a predictor of PFS. Feature 1 is linked to HOD, as a high value for $mean(\frac{\Delta C_L}{\Delta t})$ indicates increasing hyperintensity over time in the left ION and therefore the presence of HOD in the left ION. This finding suggests that an overall increase in contrast over time between the left ION tissue and surrounding brain stem tissue indicates the development of PFS, irrespective of whether the HOD is unilateral or bilateral, and whether the left or the right ION is brighter at any point throughout the patient's treatment. These findings are in keeping with the results of a recent study where damage to the right efferent cerebellar pathway, which communicates with the left ION, had a significant association with the development of PFS.[4,18,19]

The results yielded by SFS and SFFS, exhibited in table 5 and table 6, also indicate feature 1 as the most relevant feature, with all other features scoring negligible relevance scores in comparison to feature 1. Feature 1 consistently scored 68.38 or higher throughout all four tests (k =3,4,5 and 6) for all the segmentation test sets. The relevance scores for the other features in the feature set scored at least 70 times lower. This further proves the relevance of an increase in intensity of the left ION in the onset of PFS.

It should be noted that the search strategies used in this study are not optimal and are prone to local minima, with the exception of SFFS which makes an attempt at eliminating irrelevant features by carrying out a backward search in addition to the forward search. Notwithstanding this, the feature selection methods carried out in this study are ideal in a clinical scenario, more so than other methods, such as PCA, as the features retain interpretability after the feature selection techniques are applied.

The results yielded by the SVM classifier, shown in Figure 3, show an increase in classifier accuracy as the least diagnostically relevant features were eliminated. The SVM classifier reached an accuracy of 89.29%, 78.57% and 85.71%, respectively for each segmentation test set, when the only feature included was the one selected by the RSFS, SFS SFFS algorithms, that is, quantified contrast in the left ION. The area under the curve was also optimised when classification was carried out using feature 1, with values of 0.89, 0.85, and 0.89 for segmentation test sets 1,2, and 3, respectively.

From table 7, it is evident that for each segmentation test set the AUC and the accuracy is increased when only feature 1 is used. The p-value measuring the difference between the AUCs for the full feature set and the AUCs for feature 1 is statistically significant by conventional criteria. This implies that the performance of the SVM classifier is improved if only feature 1 is considered.

The average slope of contrast in the left ION is obtained by image analysis and is therefore objective, whilst the diagnosis of HOD (by radiological assessment) is made by human assessment and is subjective and prone to human error. This shows that quantified contrast in the left ION can be used a bio-marker for PFS following posterior fossa tumour resection. This is one of the pioneering studies correlating HOD and PFS using semi-automated image analysis. A previous study exists, however it did not make use of semi-automated image analysis and instead relied on human observation to identify HOD in each MRI. Such analysis is subjective and prone to human error.[4]

6 Conclusion

The aim of the experiment was to investigate the link between PFS and HOD in order to build upon the existing evidence on the development of PFS and to lead to a deeper understanding of the pathogenesis of the syndrome. A dataset of 28 patients was included in this study. The main contribution of this work consists of the quantification of HOD using automated imaging feature extraction to describe changes in intensity and size of the ION longitudinally.

This study has identified intensity, or $mean(\frac{\Delta C_L}{\Delta t})$, in the left inferior olivary nucleus (ION) as the most diagnostically relevant feature that correlates with the development of posterior fossa syndrome (PFS) following tumour resection in the posterior fossa.

Other features, including clinical features, consistently scored lower than the average slope of contrast in the left ION, throughout this study. Our findings indicate that the presence of HOD, specifically in the left ION, is highly associated with the onset of PFS following tumour resection surgery in the posterior fossa. These findings lend quantitative support to our hypothesis that there is a correlation between PFS and the occurrence of HOD following tumour resection in the posterior fossa, based on qualitative assessment of imaging. These results suggest common anatomical substrates are involved in the development of PFS and HOD and indicate an element of laterality in the development of this syndrome. This is the first study to quantify HOD using semi-automated image analysis adding reproducible and quantitative evidence to the proven hypothesis that HOD correlates with PFS.

Acknowledgments

We acknowledge the support of the Rabin Ezra Scholarship Fund in the form of a bursary, which enabled the open-access publication of this submission.

References

1 J.Siffert, T.Y.Poussaint, and L. et al., "Neurological dysfunction associated with postoperative cerebellar mutism," *Journal of Neurooncology* **48**, 75–81 (2000).

2 E.A.Kirk, V.C.Howard, and C.A.Scott, "Description of posterior fossa syndrome in children after posterior fossa brain tumor surgery ," *Journal of Pediatric Oncology Nursing* **12**(4), 181 – 187 (1995).

3 H.Baillieux, F.Weyns, and P. et al., "Posterior fossa syndrome after a vermian stroke: A new case and review of the literature," *Pediatric Neurosurgery* **43**, 386–395 (2007).

4 Z.Patay, J.Enterkin, J. Harreld, Y. Yuan, U. Ldobel, Z.Rumboldt, R. Khan, and F. Boop, "MR imaging evaluation of inferior olivary nuclei: Comparison of postoperative subjects with and without posterior fossa syndrome," *American Journal of Neuroradiology* **35** (2013).

5 MR.Chicoine, CC.Lim, JA.Evans, A.Singla, GJ.Zipfel, KM.Rich, JL.Dowling, JR.Leonard, MD.Smyth, P.Santiago, EC.Leuthardt, DD.Limbrick, and RG.Dacey, "Implementation and preliminary clinical experience with the use of ceiling mounted mobile high field intraoperative magnetic resonance imaging between two operating rooms.," *Acta Neurochir Suppl.* **109**, 97–102 (2011).

6 N.Sannai and MS.Berger, "Operative techniques for gliomas and the value of extent of resection," *Neurotherapeutics* **6**, 478–486 (2009).

7 H.Tan, "Estimation ofk-space trajectories in spiral mri," *Magnetic resonance in medicine* **61**, 1396 (2009).

8 B.M.A.Delattre, R.M.Heidemann, L.A.Crowe, J.P.Valle, and J.N.Hyacinthe, "Spiral demystified," *Magnetic Resonance Imaging* .

9 C.Rorden and M.Brett, "Stereotaxic display of brain lesions," *Behavioural Neurology* **12**, 191–200 (2000).

10 S.Karamizadeh1, S.M.Abdullah, A.A.Manaf, M.Zamani, and A.Hooman, "An overview of principal component analysis," *Journal of Signal and Information Processing* **4**, 173–175 (2013).

11 G.H.John, R.Kohavi, and K.Pfleger, "Irrelevant features and the subset selection problem," *Proceedings of the International Conference on Machine Learning* **11**, 121–129 (1994).

12 I.Kojadinovic and T.Wottka, "Comparison between a filter and a wrapper approach to variable subset selection in regression problems."

13 O. R. J. Pohjalainen and S. Kadioglu, "Feature selection methods and their combinations in high-dimensional classification of speaker likability, intelligibility and personality traits," *Computer Speech and Language* (2014).

14 A.W.Whitney, "A direct method of nonparametric measurement selection," *IEEE Transactions on Computers* **C-20**, 1100–1103 (1971).

15 O.Rasanen and J.Pohjalainen, "Random subset feature selection in automatic recognition of developmental disorders, affective states, and level of conflict from speech," *Proc. Interspeech'2013, Lyon, France* (2013).

16 C.Campbell, "Kernel methods:a survey of current techniques," *Neurocomputing* .

17 J.Shawe-Taylor and N.Cristianini, *Kernel Methods for Pattern Analysis*, Cambridge University Press, Cambridge (2004).

18 N.Law, M.Greenberg, E.Bouffet, M.D.Taylor, S.Laughlin, D.Strother, C.Fryer, D.McConnell, J.Hukin, C.Kaise, F.Wang, and D.J.Mabbott, "Clinical and neuroanatomical predictors of cerebellar mutism syndrome," *Neuro-Oncology* **14**, 1294–1303 (2012).

19 S.Kpeli, B.Yaln, B.Bilginer, N.Akalan, P.Haksal, and M.Bykpamuku, "Posterior fossa syndrome after posterior fossa surgery in children with brain tumors," *Pediatric Blood & Cancer* **56**(2), 206–210 (2011).

Michaela Spiteri is a doctoral student in Electronic Engineering at the Centre for Vision, Speech, and Signal Processing (CVSSP) at the University of Surrey, UK, under Dr.Emma Lewis. Her thesis is entitled Longitudinal MRI assessment of Brain Tumors in Children. Before coming to Surrey she completed an MSc in Biomedical Engineering (2012-2013). Before that she received a B.Eng (HONS) in Electrical and Electronic Engineering (2012) from the University of Malta. She is a member of SPIE.

David Windridge is a Senior Lecturer in Computer Science at Middlesex University and heads the university's Data Science activities. He is a Visiting Professor at Trento University, Italy, and a visiting Senior Research Fellow at the University of Surrey (he was previously a Senior Research Fellow within the Centre for Vision, Speech and Signal Processing). His research interests centre on the fields of machine learning (particularly kernel methods classifier-fusion), cognitive systems and computer vision. He has authored more than 80 peer-reviewed publications.

Shivaram Avula is a Consultant Paediatric Radiologist at Alder Hey with an interest in Neuroimaging. He completed Paediatric Radiology fellowships at Alder Hey and the Hospital for Sick Children in Toronto, Canada. His clinical interests include head neck imaging, neuroimaging for brain tumours, epilepsy and other childhood neurological disorders. Research interests includes advanced neuro MRI techniques for brain tumours and intraoperative MRI .

Ram Kumar has been a Consultant Paediatric Neurologist at Alder Hey Childrens Hospital in Liverpool since early 2007. He is the Service Group Lead for Neurology, Neurosurgery, Long-Term Ventilation, Rehabilitation and Sleep services. He obtained an MA in Pathology from the University of Cambridge in 1992, a Bachelor of Medicine and Surgery from the same university in 1994 and an MRCP(UK) in 1998. He joined the GMC Specialist Register CCT in Paediatrics and Paediatric Neurology in 2007.

Emma Lewis is a Lecturer in Medical Image Analysis at the Centre for Vision Speech and Signal Processing at the University of Surrey. Previously she was Principal Research Fellow at the Dementia Research Centre, at The National Hospital for Neurology and Neurosurgery, London. Her research involves applying image processing techniques within a variety of clinical application areas and image modalities. In particular, Emma has developed methods for quantification, simulation and artifact correction within domains including neuroimaging, mammography and nuclear medicine motion correction.

List of Figures

List of Tables

5 The average relevance scores calculated by the sequential forward selection algorithm over 100 iterations

6 The average relevance scores calculated by the sequential floating forward selection algorithm over 100 iterations

7 Area under the curve and accuracy for the SVM classifier used on: the full feature dataset and feature 1 on Test sets 1,2, and 3