Improved prediction of postoperative paediatric cerebellar mutism syndrome using an artificial neural network

Jai Sidpra[1,2,3], Adam P. Marcus[4], Ulrike Löbel[3], Sebastian M. Toescu[5,6], Derek Yecies[7], Gerald Grant[7], Kristen Yeom[8], David M. Mirsky[9], Hani J. Marcus[10,11], Kristian Aquilina[2,5], Kshitij Mankad[2,3]

[1] University College London Medical School, London, WC1E 6DE, UK (J.S.)

[2] Developmental Biology and Cancer Section, University College London Great Ormond Street Institute of Child Health, London, WC1N 1EH, UK (J.S., K.A., and K.M.)

[3] Department of Neuroradiology, Great Ormond Street Hospital for Children NHS Foundation Trust, London, WC1N 3JH, UK (J.S., U.L., and K.M.)

[4] Department of Brain Sciences and Computing, Imperial College London, London, SW7 2BU, UK (A.P.M.)

[5] Department of Neurosurgery, Great Ormond Street Hospital for Children NHS Foundation Trust, London, WC1N 3JH, UK (S.M.T. and K.A.)

[6] Developmental Imaging and Biophysics Section, University College London Great Ormond Street Institute of Child Health, London WC1N 1EH, UK (S.M.T.)

[7] Department of Neurosurgery, Lucile Packard Children's Hospital, Stanford, CA 94304, USA (D.Y. and G.G.)

[8] Department of Neuroradiology, Lucile Packard Children's Hospital, Stanford, CA 94304, USA (K.Y.)

[9] Department of Radiology, Children's Hospital Colorado, University of Colorado School of Medicine, Aurora, CO 80045, USA (D.M.M.)

[10] Department of Neurosurgery, National Hospital for Neurology and Neurosurgery, London, WC1N 3BG, UK (H.J.M.)

[11] Wellcome / EPSRC Centre for Interventional and Surgical Sciences, University College London, London, WC1E 6BT, UK (H.J.M.)

Corresponding Author Contact Information

Dr Kshitij Mankad

Department of Neuroradiology, Great Ormond Street Hospital for Children NHS Foundation Trust, Great Ormond Street, London, WC1N 3JH, UK

E-Mail: drmankad@gmail.com

Mobile: +44 (0)7861 639 010

Author Contributions

Study conception: J.S., H.J.M, K.A., and K.M.

Data acquisition: J.S., U.L., S.M.T, D.Y., G.G., K.Y., D.M.M., and K.M.

Data analysis: J.S., A.P.M., and H.J.M.

Manuscript draft: J.S.

Manuscript revisions: J.S., A.P.M., U.L., S.M.T, D.Y., G.G., K.Y., D.M.M., H.J.M., K.A., and K.M.

Acknowledgements

We would like to thank Dr Olivia Carney for her invaluable critique of our methods. We would also like to acknowledge the neuro-oncology multidisciplinary teams at Great Ormond Street Hospital for Children, Children's Hospital Colorado, and Lucile Packard Children's Hospital and the steadfast work they do for the benefit of children such as these.

Compliance with ethical standards

Funding

No funding was sought for this work. A.P.M. is funded by the United Kingdom Research and Innovation Centre for Doctoral Training in Artificial Intelligence for Healthcare (grant number: P/S023283/1). S.M.T. is funded by the Great Ormond Street Hospital Children's Charity and is an Honorary Research Fellow of the Royal College of Surgeons of England. H.J.M. is funded by the Wellcome / EPSRC Centre for Interventional and Surgical Sciences and the National Institute of Health Research University College London Biomedical Research Centre. K.M. conducts medicolegal expert work and receives speaker honoraria from Guerbet, Novartis, and Siemens.

All research at Great Ormond Street Hospital for Children NHS Foundation Trust and the University College London Great Ormond Street Institute of Child Health is made possible by the Great Ormond Street National Institute for Health Research Biomedical Research Centre. The views expressed are those of the authors and not necessarily those of the National Health Service, the National Institute for Health Research, or the Department of Health.

Conflict of interest

On behalf of all authors, the corresponding author asserts that no financial relationships exist with any organisations that might have an interest in the submitted work and that no other relationships or activities exist that could appear to have influenced the submitted work.

Ethical Approval

Ethical approval was obtained prior to the commencement of this study from: (i) the Department of Research Management and Governance, Great Ormond Street Hospital for Children NHS Foundation Trust (reference numbers: 17NI17, 19NI01, 19NI07); (ii) the Colorado Multiple Institutional Review Board (reference number: 21-3146); and (iii) the Research Compliance Office of Stanford University. At all institutions, the need for informed consent was waived for this retrospective study.

Data Availability Statement

Anonymised data is available upon request.

Code Availability Statement

Code is available from the following GitHub repository:

https://github.com/amarcu5/cerebellar-mutism-prediction

Our online calculator is available from the following GitHub page:

https://amarcu5.github.io/cerebellar-mutism-prediction/calc.htm

Abstract

Background:

Postoperative paediatric cerebellar mutism syndrome (pCMS) is a common but severe complication which may arise following the resection of posterior fossa tumours in children. Two previous studies have aimed to preoperatively predict pCMS, with varying results. In this work, we examine the generalisation of these models and determine if pCMS can be predicted more accurately using an artificial neural network (ANN).

Methods:

An overview of reviews was performed to identify risk factors for pCMS, and a retrospective dataset collected as per these defined risk factors from children undergoing resection of primary posterior fossa tumours. The ANN was trained on this dataset and its performance evaluated in comparison to logistic regression and other predictive indices via analysis of receiver operator characteristic curves. Area under the curve (AUC) and accuracy were calculated and compared using a Wilcoxon signed rank test, with $p<0.05$ considered statistically significant.

Results:

204 children were included, of whom 80 developed pCMS. The performance of the ANN (AUC 0.949; accuracy 90.9%) exceeded that of logistic regression ($p<0.05$) and both external models ($p<0.001$).

Conclusion:

Using an ANN, we show improved prediction of pCMS in comparison to previous models and conventional methods.

Key words:

Artificial neural network; complications; magnetic resonance imaging; post-operative paediatric cerebellar mutism syndrome; posterior fossa tumour.

Key Points

1. We identify predictive and reproducible anatomical risk factors of cerebellar mutism syndrome.
2. Cerebellar mutism syndrome can be accurately predicted using an artificial neural network.

Importance of the study

Postoperative paediatric cerebellar mutism syndrome (pCMS) is a severe complication of childhood posterior fossa tumour resection. Despite its relatively high incidence, the syndrome's aetiology and predisposing risk factors are poorly understood. As such, not only is the prediction of pCMS clinically challenging but this knowledge gap represents a significant area of uncertainty for patients giving consent prior to surgery and for clinicians developing postoperative management plans. Our study, the first to implement an artificial neural network for the prediction of complications in paediatric neuro-oncology, demonstrates that pCMS can be accurately and reproducibly predicted. By interpreting logistic regression coefficients, we also identify the neuroanatomical features most predictive of pCMS. It is our hope that this model will facilitate the routine clinical prediction of pCMS and act as a novel clinical decision making tool: permitting the consideration of less aggressive resective surgery in combination with adjuvant therapies for patients at greatest risk.

Introduction

Brain tumours are the most common solid tumours of childhood and the leading cause of cancer-related deaths in children.[1] The vast majority (60-70%) arise within the posterior fossa, an area dense with eloquent neural parenchyma, and have an outcome dependent upon the extent of surgical resection.[2,3] As such, the neurosurgeon must take care to balance the prognostic benefits of decompression and cytoreduction with the risk of incurring long-lasting neurological sequelae.

Despite recent advances in surgical management, perioperative imaging, and adjuvant therapies, postoperative paediatric cerebellar mutism syndrome (pCMS) remains an enduring complication of posterior fossa tumour resection, with a reported prevalence of 8-39%.[4,5] Characterised by a delayed onset transient mutism, emotional lability, ataxia, and hypotonia, pCMS is thought to arise secondary to proximal damage to the efferent cerebellar pathways during surgery.[6,7] Children typically recover slowly over several months, though their speech may never return to normal, and they may be late to reach developmental milestones – requiring intensive physiotherapy and speech and language therapy throughout the course of their recovery.[8,9] Long-term neuropsychological deficits have also been reported in children with pCMS, with sparse recovery and the potential for further neurocognitive decline during development.[10]

Machine learning is a form of artificial intelligence in which algorithms improve their performance with experience (training). Given their ability to identify complex patterns in large datasets, such computational methods have the potential to improve outcome prediction beyond an individual surgeon's clinical intuition.[11] In light of the severity and long-term impact of pCMS, the accurate prediction of its onset will enable clinicians to counsel families preoperatively, as well as to formulate postoperative management plans, and ultimately, to implement risk-mitigating measures towards the prevention of pCMS, with consideration of less aggressive resective surgery for children at greatest risk.

Two previous studies have attempted to preoperatively predict pCMS using machine learning.[12,13] However, these are trained on small datasets; employ quite restricted models which permit little interaction between input risk factors; and lack external validation. Our aim is to interrogate the external validity of these models and to improve the prediction of pCMS by training and validating an artificial neural network (ANN) on a large series of operated children with primary posterior fossa tumours. We hypothesised that an ANN would improve the prediction of pCMS as it considers implicit and complex non-linear interactions between input risk factors. Indeed, the efficacy of this learning framework on relatively small datasets within neurosurgery has recently been shown.[14,15]

Materials and methods

This study was approved by the institutional review boards of all collaborating institutions prior to commencement.

I. Defining risk factors for pCMS

An overview of reviews was performed in accordance with The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement to identify risk factors for pCMS.[16]

Inclusion criteria

We included review articles written in the English language that report clinical and radiological features predictive of pCMS in children (≤18 years).

Databases, search strategy, and study selection

Ovid MEDLINE and EmBase were searched using the following strategy: ((*akinetic mutism*) OR (*cerebellar cognitive and affective syndrome*) OR (CCAS) OR (*Schmahmann's syndrome*) OR (*posterior fossa syndrome*) OR (*transient cerebellar mutism*) OR (*transient cerebellar mutism and subsequent dysarthria*) OR (*cerebellar mutism*) OR (*cerebellar mutism syndrome*) OR (*post-operative paediatric cerebellar mutism syndrome*) OR (*post-operative pediatric cerebellar mutism syndrome*) OR (*POPCMS*) OR (*POP-CMS*) OR (*CMS*) OR (*pCMS*)).

Titles and abstracts were screened to identify articles which met the inclusion criteria (Supplementary Figure 1). Full text review of these articles was then performed, and reported risk factors recorded (Supplementary Table 1).[5,8,9,17-62]

Refinement of risk factors

Several risk factors were excluded prior to data collection: patient socioeconomic status, handedness, and preoperative language and/or behavioural impairment could not be reliably determined from retrospective medical records. Surgical risk factors such as the extent of resection, surgical approach, and use of a cavitron ultrasonic surgical aspirator were also excluded, as they are determined intraoperatively rather than preoperatively and so cannot be reliably included in a preoperative risk prediction model.

Features did not have to reach a certain risk threshold to be included. Given our relatively poor understanding of the pathophysiology underlying pCMS, a key advantage of using an ANN for prediction is that the model itself selects and weights the most predictive features.

II. Patient selection and data curation

The Strengthening Reporting of Observational Studies in Epidemiology (STROBE) statement was used to inform the collation of our retrospective cohort.[63] Modelling methods and results are reported in line with the Transparent Reporting of A Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) statement.[64]

A prospectively maintained neuro-oncology database was searched for all patients with primary posterior fossa tumours who underwent craniotomy and resection at our institution between 1 January 2002 and 1 January 2021. All patients received treatment as per local guidelines including case discussion within a dedicated paediatric neuro-oncology multidisciplinary team.

Clinical notes were interrogated for demographics and relevant surgical risk factors. pCMS was diagnosed in patients meeting the 2016 consensus criteria: transient reduced speech or mutism and emotional lability of delayed onset following resection of their posterior fossa tumour.[43] Preoperative magnetic resonance imaging (MRI) of the neuraxis was reviewed independently by three observers (J.S., U.L., and K.M.) blinded to outcome, with imaging features and measurements recorded as per our defined risk factors. All imaging was obtained electronically as part of routine clinical care and did not undergo any subsequent modification. Compression, signal change, and infiltration were assessed on all sequences and confirmed on two planes. Interobserver agreement was calculated for qualitative variables using Fleiss's κ, with the minimum cut-off for inclusion defined as $\kappa \geq 0.60$, and for quantitative variables, using the intraclass correlation coefficient with the same inclusion cut-off.[65] Table 1 defines the variables we selected for data collection in addition to the interobserver reliability achieved. Due to their low interobserver agreement, risk factors involving the dentate, red, and inferior olivary nuclei were not included in the model. All other variables were included in both the ANN and logistic regression models.

To define the final inputs of our model, we used the modal value for qualitative variables and the mean value for quantitative variables. Missing data was encountered at random in two instances and was handled in the following manner:

1. Children without apparent diffusion coefficient (ADC) values on preoperative MRI. The ADC represents a quantitative indicator of the diffusion of water molecules within a tissue and has well-reported ranges in the literature corresponding to tumour type.[67] Hence, the mean ADC was calculated for each tumour type in our cohort for children with and without pCMS. This mean ADC was then used in place of missing values (n=36).
2. Children who underwent cerebrospinal fluid diversion prior to preoperative MRI and tumour resection. In these cases, as children has reached the threshold for neurosurgical intervention, they were assumed to have a hydrocephalus severity score of three and their Evan's index was imputed as the mean of children with this score who did and did not develop pCMS (n=12).

In order to increase the number of patients included with pCMS and, therefore, the potential accuracy of our model, additional patients with pCMS were included from two collaborating institutions subject to identical inclusion criteria. Preoperative imaging from these institutions was read in consensus between J.S., K.M., and the collaborating author (S.M.T. or D.M.M.). As a high interobserver reliability was achieved across three observers on our large single centre cohort, and external data constituted less than 10% of all included patients, a consensus agreement between three authors was deemed sufficient for the inclusion of this data. In areas of disagreement, the modal value was taken (i.e., agreement between a minimum of two of the three observers).

III. ANN theory

We implement a previously reported ANN with proven accuracy within neurosurgery.[15] A complete description of our computational approach is presented in the Supplementary Materials.

In brief, we used nested cross-validation: the inner loop performed evolutionary hyperparameter optimisation across 100 repeats of 10-fold stratified cross-validation whilst the outer loop evaluated the network across 10 repeats of 10-fold stratified cross-validation. The final model was a stacked ensemble of 1000 constituent ANNs producing a single output. The loss function minimised during training was the mean squared error between the ANN prediction and the patient's clinical outcome. The fitness of each solution was defined by the average validation error of the set of ANNs. Early stopping was applied to improve generalisation and to prevent overfitting.

A cross-validated paired Wilcoxon signed rank test was used to interrogate the statistical significance of the difference in performance between (1) our ANN and logistic regression and (2) our ANN and the Liu et al. and Dhaenens et al. models. Averaged receiver operating characteristic (ROC) curves were created for these comparative models using 10 repeats of 10-fold stratified cross-validation (the outer loop of nested cross-validation). The accuracy of the final models was determined by comparing the ANN prediction with the patient's clinical outcome. The area under the curve (AUC) was calculated for each of the ROC curves and evaluated to compare the discriminatory power of the models. To evaluate the fit of the models, sensitivity, specificity, negative and positive predictive values were calculated using the same approach. In all instances, $p<0.05$ was considered statistically significant.

Results

Patient sample characteristics

Retrospective review of medical records identified 426 children who underwent resection of primary posterior fossa tumours at our institution, of whom 66 (15.5%) developed pCMS. Four patients with pCMS were excluded due to lack of preoperative MRI. Eighteen patients with pCMS were added from collaborating centres. Our final dataset consisted of all pCMS patients with pre-operative MRI (n=80) in addition to 124 patients without pCMS randomly sampled from the remaining dataset via a random number generator. The decision to train our model on a subsample of 124 patients who did not develop pCMS rather than the total 360 was pragmatic and constraint-based. Indeed, given the high reported accuracy of our ANN as a neurosurgical predictive model when trained on a dataset of 135 patients, training on this large subset should not negatively impact the performance of the ANN.[15]

Patient demographics, tumour characteristics, and surgical management are summarised in Table 2.

Network selection

The mean network structure had three layers: a 55-neuron input layer of risk factors for pCMS; one hidden layer containing eight neurons; and an output layer containing one neuron representing the probability of a patient developing pCMS.

The optimal mean network hyperparameters defined by evolutionary hyperparameter optimisation are detailed in Supplementary Table 2.

Network evaluation

The comparative performance of the ANN is reported in Figure 1 and is illustrated as the mean ROC curve.

The AUC and accuracy of the ANN exceeded that of any other model ($p<0.05$ vs. logistic regression; $p<0.001$ vs. Liu *et al*. and Dhaenens *et al*.). The ANN also outperformed all models in terms of its sensitivity and negative predictive value ($p<0.001$ in all instances). Though the specificity and positive predictive value of the ANN were not significantly different to logistic regression, and worse than the Dhaenens *et al*. model, the ANN's more substantive improvements in sensitivity and negative predictive value rendered it a more accurate predictive test for pCMS.

The most predictive neuroanatomical features of pCMS as determined by logistic regression coefficients are illustrated in Figure 2. Notably, tumoral involvement of structures associated with the dentatorubrothalamocortical tracts is shown to heighten the risk of developing pCMS, thereby lending further support to the implication of this tract in the pathophysiology of pCMS.[9,20]

Discussion

This is the first study to report the use of an ANN for the prediction of complications in paediatric neuro-oncology. This is also the first study to implement an ANN for the prediction of pCMS and to demonstrate improved accuracy of the ANN over and above existing risk prediction models. Though the clinical significance of this improved accuracy in comparison to logistic regression remains uncertain, given the severity of pCMS and the relatively high volume of posterior fossa tumours encountered within paediatric neurosurgical practice, the increase in accuracy achieved may endow some benefit.

It is our hope that the model (available as an online calculator: https://amarcu5.github.io/cerebellar-mutism-prediction/calc.htm) will act as a useful adjunct to surgical decision making and as a counselling tool for children and their families when giving informed consent prior to surgery. Figure 3 highlights the potential utility of our model in clinical practice by illustrating ANN predictions for three patients: one who developed pCMS and two who did not.

Since 1990, survival rates for children with medulloblastoma, the most common tumour type implicated in pCMS, have remained relatively constant, in part due to the increasing importance of balancing the side-effects of aggressive therapy (such as pCMS) with the potential for improved outcome.[68] A large study of 787 children with medulloblastoma has shown that, though gross total resection remains the gold-standard surgical outcome, children left with minimal residual tumour following subtotal resection can expect similar outcomes.[69] Hence, in the context of continuing uncertainty as to the oncological importance of complete resection in medulloblastoma, an accurate predictive model such as ours may permit the development of a clinical trial in which limited initial surgical resection, followed by adjuvant therapy and late second-look surgery of a smaller tumour, may be considered for children at very high risk of developing pCMS.

Liu *et al*. first aimed to predict pCMS using a C4.5 decision tree and reported an accuracy of 88.8%.[12] Subsequent attempts to validate this model by both our study and Dhaenens *et al*. have been unsuccessful, with reported accuracies of 39.8% and 78%, respectively.[12,13] This poor generalisation performance may be partly explained by Liu *et al*.'s model architecture: a single, non-ensemble decision tree which is consequently more susceptible to noise and has the potential to overfit. Aiming to improve upon Liu *et al*.'s work, Dhaenens *et al*. implemented a logistic regression model and reported an accuracy of 87%, which generalises well to our cohort (accuracy 88.7%).[13] Indeed, the comparatively weaker classification performance of decision trees when compared to logistic regression and ANNs has been shown empirically.[70]

The improved accuracy of the ANN (90.9%) most likely lies in its ability to weigh complex non-linear relationships between variables, such as those underlying pCMS, when little is known about their distribution and interaction. Hence, through this work, we also reinforce the fact that accurate ANNs can be developed on relatively small datasets by following established best practices: using a stacked ensemble; taking the mean performance of multiple runs; and evaluating the model using k-fold cross-validation.[71-73]

This work does, however, have several limitations. First, we employed a retrospective study design which rendered several risk factors indeterminable and risked the introduction of selection bias. However, given the relative rarity of paediatric brain tumours, this choice was pragmatic and enabled us to report one of the largest cohorts of children with pCMS in the literature. By sampling children with pCMS from multiple centres, we also increased the replicability of our model beyond a single institution. Second, we predominantly trained our ANN on imaging features predictive of pCMS from preoperative MRI. Whilst this did permit direct comparison with and interrogation of other models to date, we hypothesise that the accuracy of our model would be improved by training a convolutional neural network on raw preoperative imaging. This shift towards automated prediction would also increase the practical utility of our model by eliminating the time-consuming process of image interpretation and manual measurement. Third, due to the nature of preoperative predictions, we are unable to account for individual variation between surgeons, approaches, and techniques employed. Most notably, all surgeons and institutions will differ in how they balance the benefits of more aggressive surgery, and a potential gross total resection, with the risk of incurring pCMS.

For the above reasons, our results should not be taken as direct clinical recommendations at this stage. However, we will aim to address these limitations in a planned prospective multicentre study. We will also aim to expand our model to consider and predict other common postoperative complications of posterior fossa tumour resection in children, namely: disturbances of motor function and gait; cranial nerve deficits; and visual impairment.[4] A prospective study would also permit the integration and analysis of recently identified surgical risk factors excluded in this retrospective study, namely surgical experience and extent of resection.[74,75]

Conclusion

We present a novel framework that interprets features from preoperative MRI and more accurately predicts the likelihood of a patient developing pCMS than previous methods. It is our hope that, following prospective validation of our model, the routine clinical prediction of pCMS will lead to safer surgery and better-informed discussions of the risks involved with patients and their families. As such, this work represents an exciting step towards the personalised, risk-stratified surgical management of children with brain tumours.

References

1. Pollack IF, Agnihotri S, Broniscer A. Childhood brain tumors: current management, biological insights, and future directions. *J Neurosurg. Pediatrics*. 2019;23:261-273.
2. Dolecek TA, Propp JM, Stroup NE, Kruchko C. CBTRUS statistical report: primary brain and central nervous system tumours diagnosed in the United States in 2005-2009. *Neuro-Oncology*. 2012;14(Supplement 5):v1-v49.
3. Schneider CA, Shimony N, Kothbauer KF. Posterior fossa tumours. In: Jallo GI, Kothbauer KF, Recinos VM, eds. *Handbook of Pediatric Neurosurgery*. Thieme; 2018:154-168.
4. Toescu SM, Samarth G, Horsfall HL, *et al*. Fourth ventricle tumors in children: complications and influence of surgical approach. *J Neurosurg. Pediatrics*. 2021;27:52-61.
5. Pitsika M, Tsitouras V. Cerebellar Mutism: A Review. *J Neurosurg. Pediatrics*. 2013;12:604-614.
6. Miller NG, Reddick WE, Kocak M, *et al*. Cerebellocerebral diaschisis is the likely mechanism of postsurgical posterior fossa syndrome in pediatric patients with midline cerebellar tumors. *AJNR Am J Neuroradiol*. 2010;31(2):288-94.
7. Patay Z, Enterkin J, Harreld JH, *et al*. MR imaging evaluation of inferior olivary nuclei: comparison of postoperative subjects with and without posterior fossa syndrome. *AJNR Am J Neuroradiol*. 2014 Apr;35(4):797-802.
8. Catsman-Berrevoets C, Patay Z. Cerebellar mutism syndrome. In: Manto M, Huisman TAGM, eds. *Hand Clin Neurol*. Elsevier; 2018;155:273-288.
9. Patay Z. Postoperative posterior fossa syndrome: unraveling the etiology and underlying pathophysiology by using magnetic resonance imaging. *Childs Nerv Syst*. 2015;31:1853-1858.
10. Schreiber JE, Palmer SL, Conklin HM, *et al*. Posterior fossa syndrome and long-term neuropsychological outcomes among children treated for medulloblastoma on a multi-institutional, prospective study. *Neuro-Oncology*. 2017;19(12):1673-1682.

11. Nagendran M, Chen Y, Lovejoy CA, *et al*. Artificial intelligence versus clinicians: systematic review of design, reporting standards, and claims of deep learning studies. *BMJ*. 2020 Mar 25;368:m689.
12. Liu JF, Dineen RA, Avula S, *et al*. Development of a pre-operative scoring system for predicting risk of post-operative paediatric cerebellar mutism syndrome. *Br J Neurosurg*. 2018;32(1):18-27.
13. Dhaenens BAE, Van Veelen MLC, Catsman-Berrevoets CE. Preoperative prediction of postoperative cerebellar mutism syndrome. Validation of existing MRI models and proposal of the new Rotterdam pCMS prediction model. *Childs Nerv Syst*. 2020;36:1471-1480.
14. Marcus HJ, Williams S, Hughes-Hallet A, Camp SJ, Nandi D, Thorne L. Predicting surgical outcome in patients with glioblastoma multiforme using pre-operative magnetic resonance imaging: development and preliminary validation of a grading system *Neurosurg Rev*. 2017;40:621-631.
15. Marcus AP, Marcus HJ, Camp SJ, Nandi D, Kitchen N, Thorne L. Improved prediction of surgical resectability in patients with glioblastoma using an artificial neural network. *Sci Rep*. 2020;10:5143.
16. Page MJ, McKenzie JE, Bossuyt PM, *et al*. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*. 2021;372:n71.
17. Toescu SM, Hales PW, Aquilina K, Clark CA. Quantitative MRI in post-operative paediatric cerebellar mutism syndrome. *Eur J Radiol*. 2018;108:43-51.
18. Catsman-Berrevoets CE. Cerebellar mutism syndrome: cause and rehabilitation. *Curr Opin Neurol*. 2017;30(2):133-139.
19. Avula S, Mallucci C, Kumar R, Pizer B. Posterior fossa syndrome following brain tumour resection: review of pathophysiology and a new hypothesis on its pathogenesis. *Childs Nerv Syst*. 2015;31(10):1859-67.
20. Reed-Berendt R, Phillips B, Picton S, *et al*. Cause and outcome of cerebellar mutism: evidence from a systematic review. *Childs Nerv Syst*. 2014;30(3):375-85.
21. Gudrunardottir T, Sehested A, Juhler M, Schmiegelow K. Cerebellar mutism: review of the literature. *Childs Nerv Syst*. 2011;27(3):355-63.

22. Wells EM, Walsh KS, Khademian ZP, Keating RF, Packer RJ. The cerebellar mutism syndrome and its relation to cerebellar cognitive function and the cerebellar cognitive affective disorder. *Dev Disabil Res Rev*. 2008;14(3):221-8.
23. Charalambides C, Dinopoulos A, Sgouros S. Neuropsychological sequelae and quality of life following treatment of posterior fossa ependymomas in children. *Childs Nerv Syst*. 2009;25(10):1313-20.
24. Ozgur BM, Berberian J, Aryan HE, Meltzer HS, Levy ML. The pathophysiologic mechanism of cerebellar mutism. *Surg Neurol*. 2006;66(1):18-25.
25. Ersahin Y, Mutluer S, Cagli S, Duman Y. Cerebellar mutism: report of seven cases and review of the literature. *Neurosurgery*. 1996;38(1):60-5;discussion 66.
26. Turgut M. Transient "cerebellar" mutism. *Childs Nerv Syst*. 1998;14(4-5):161-6.
27. Gelabert-Gonzalez M, Fernandez-Villa J. Mutism after posterior fossa surgery. Review of the literature. *Clin Neurol Neurosurg*. 2001;103(2):111-4.
28. Ildan F, Tuna M, Erman T, Gocer AI, Zeren M, Cetinalp E. The evaluation and comparison of cerebellar mutism in children and adults after posterior fossa surgery: report of two adult cases and review of the literature. *Acta Neurochir (Wien)*. 2002;144(5):463-73.
29. Giacino JT, Fins JJ, Laureys S, Schiff ND. Disorders of consciousness after acquired brain injury: The state of the science. *Nat. Rev. Neurol*. 2014;10(2):99-114.
30. Pollack IF. Neurobehavioral abnormalities after posterior fossa surgery in children. *Int. Rev. Psychiatry*. 2001;13(4):302-312.
31. De Smet HJ, Baillieux H, Catsman-Berrevoets C, De Deyn PP, Marien P, Paquier PF. Postoperative motor speech production in children with the syndrome of 'cerebellar' mutism and subsequent dysarthria: a critical review of the literature. *Europ J Paediatr Neurol*. 2007;11(4):193-207.
32. Aguiar PH, Plese JP, Ciquini O, Marino R. Transient mutism following a posterior fossa approach to cerebellar tumors in children: a critical review of the literature. *Childs Nerv Syst*. 1995;11(5):306-10.
33. Chevignard M, Camara-Costa H, Doz F, Dellatolas G. Core deficits and quality of survival after childhood medulloblastoma: a review. *Neurooncol Pract*. 2017;4(2):82-97.

34. Lanier JC, Abrams AN. Posterior fossa syndrome: Review of the behavioral and emotional aspects in pediatric cancer patients. *Cancer*. 2017;123(4):551-559.
35. Zyrianova Y, Alexander L, Faruqui R. Neuropsychiatric presentations and outcomes in children and adolescents with primary brain tumours: Systematic review. *Brain Inj*. 2016;30(1):1-9.
36. Manto M. Cerebellar motor syndrome from children to the elderly. *Handb. clin. neurol*. 2018;154:151-166.
37. Kuper M, Timmann D. Cerebellar mutism. *Brain Lang*. 2013;127(3):327-33.
38. Gudrunardottir T, Lannering B, Remke M, *et al*. Treatment developments and the unfolding of the quality of life discussion in childhood medulloblastoma: a review. *Childs Nerv Syst*. 2014;30(6):979-90.
39. Tasdemiroglu E, Kaya M, Yildirim CH, Firat L. Postoperative cerebellar mutism and autistic spectrum disorder. *Childs Nerv Syst*. 2011;27(6):869-78.
40. Gordon N. The cerebellum and cognition. *Europ J Paediatr Neurol*. 2007;11(4):232-4.
41. Salvati M, Missori P, Lunardi P, Orlando ER. Transient cerebellar mutism after posterior cranial fossa surgery in an adult. Case report and review of the literature. *Clin Neurol Neurosurg*. 1991;93(4):313-6.
42. Gordon N. Mutism: elective or selective, and acquired. *Brain Dev*. 2001;23(2):83-7.
43. Gudrunardottir T, Morgan AT, Lux AL, *et al*. Consensus paper on post-operative pediatric cerebellar mutism syndrome: the Iceland Delphi results *Childs Nerv Syst*. 2016;32:1195–1203.
44. Srinivasan VM, Ghali MG, North RY, Boghani Z, Hansen D, Lam S. Modern management of medulloblastoma: Molecular classification, outcomes, and the role of surgery. *Surg Neurol Int*. 2016;7(Suppl 44):S1135-S1141.
45. Beez T, Munoz-Bendix C, Steiger HJ, Hanggi D. Functional tracts of the cerebellum - essentials for the neurosurgeon. *Neurosurg Rev*. 2021;44:273–278.
46. Sankey EW, Srinivasan ES, Mehta VA, *et al*. Perioperative Assessment of Cerebellar Masses and the Potential for Cerebellar Cognitive Affective Syndrome. *World Neurosurg*. 2020;144:222-230.

47. Tamburrini G, Frassanito P, Chieffo D, Massimi L, Caldarelli M, Di Rocco C. Cerebellar mutism. *Childs Nerv Syst*. 2015;31(10):1841-51.
48. Dietze DD, Mickle JP. Cerebellar mutism after posterior fossa surgery. *Pediatr Neurosurg*. 1990-1991;16(1):25-31; discussion 31.
49. Packer RJ. Childhood medulloblastoma: progress and future challenges. *Brain Dev*. 1999;21(2):75-81.
50. Schmahmann JD. The cerebellum and cognition. *Neurosci Lett*. 2019;688:62-75.
51. Schmahmann JD, Guell X, Stoodley CJ, Halko MA. The Theory and Neuroscience of Cerebellar Cognition. *Annu Rev Neurosci*. 2019;42:337-364.
52. Argyropoulos GPD, van Dun K, Adamaszek M, *et al*. The Cerebellar Cognitive Affective/Schmahmann Syndrome: A Task Force Paper. *Cerebellum*. 2020;19(1):102-125.
53. Ghali MGZ. Telovelar surgical approach. *Neurosurg Rev*. 2021;44(1):61-76.
54. Stoodley CJ, Schmahmann JD. Functional topography of the human cerebellum. *Handb. clin. neurol*. 2018;154:59-70.
55. De Smet HJ, Paquier P, Verhoeven J, Marien P. The cerebellum: its role in language and related cognitive and affective functions. *Brain Lang*. 2013;127(3):334-42.
56. Yildiz O, Kabatas S, Yilmaz C, Altinors N, Agaoglu B. Cerebellar mutism syndrome and its relation to cerebellar cognitive and affective function: Review of the literature. *Ann. Indian Acad. Neurol*. 2010;13(1):23-27.
57. Gordon N. Speech, language, and the cerebellum. *Eur. J. Disord. Commun*. 1996;31(4):359-367.
58. Sun LR, Cooper S. Neurological Complications of the Treatment of Pediatric Neoplastic Disorders. *Pediatr Neurol*. 2018;85:33-42.
59. Hampson DR, Blatt GJ. Autism spectrum disorders and neuropathology of the cerebellum. *Front. neurosci*. 2015;9:420.
60. Manto M, Habas C. Cerebellar disorders: clinical/radiologic findings and modern imaging tools. *Handb. clin. neurol*. 2016;135:479-491.
61. Winkler EA, Birk H, Safaee M, Yue JK, *et al*. Surgical resection of fourth ventricular ependymomas: case series and technical nuances. *J Neurooncol*. 2016;130(2):341-349.

62. Shimoji K, Miyajima M, Karagiozov K, Yatomi K, Matsushima T, Arai H. Surgical considerations in fourth ventricular ependymoma with the transcerebellomedullary fissure approach in focus. *Child's Nerv. Syst*. 2009;25(10):1221-1228.
63. Von Elm E, Altman DG, Egger M, *et al*. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *BMJ*. 2007;335(7624):806–808.
64. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMJ*. 2015 Jan 7;350:g7594.
65. Landis JT, Koch GG. The Measurement of Observer Agreement for Categorical Data. *International Biometric Society*. 1977;33(1):159-174.
66. Zhang H, Liao Z, Hao X, *et al*. Establishing reproducible predictors of cerebellar mutism syndrome based on pre-operative imaging. *Childs Nerv Syst*. 2019;35:795-800.
67. Rumboldt Z, Camacho DL, Lake D, Welsh CT, Castillo M. Apparent diffusion coefficients for differentiation of cerebellar tumors in children. *AJNR Am J Neuroradiol*. 2006;27(6):1362-1369.
68. Johnston DL, Keene D, Kostova M, *et al*. Survival of children with medulloblastoma in Canada diagnosed between 1990 and 2009 inclusive. *J Neurooncol*. 2015 Sep;124(2):247-53.
69. Thompson EM, Hielscher T, Bouffet E, *et al*. Prognostic value of medulloblastoma extent of resection after accounting for molecular subgroup: a retrospective integrated clinical and molecular analysis. *Lancet Oncol*. 2016 Apr;17(4):484-495.
70. Curram S, Mingers J. Neural Networks, Decision Tree Induction and Discriminant Analysis: An Empirical Comparison. *The Journal of the Operational Research Society*. 1994;45(4):440-450.73.
71. Fushiki T. Estimation of prediction error by using K-fold cross-validation. *Stat Comput*. 2011;21:137-146.
72. Opitz DW, Maclin R. Popular ensemble methods: an empirical study. *J. Artif. Int. Res*. 1999;11:169–198.

73. Hashem S, Schmeiser B, and Yih Y. 1994. Optimal linear combinations of neural networks: An overview. In Proceedings of the 1994 IEEE International Conference on Neural Networks, Orlando, FL.

74. Khan RB, Patay Z, Klimo P, *et al*. Clinical features, neurologic recovery, and risk factors of postoperative posterior fossa syndrome and delayed recovery: a prospective study. *Neuro-Oncology*. 2021;23(9):1586-1596.

75. Grønbæk JK, Wibroe M, Toescu S, *et al*. Postoperative speech impairment and surgical approach to posterior fossa tumours in children: a prospective European multicentre cohort study. *Lancet Child Adolesc Health*. 2021; S2352-4642(21)00274-1.

Figure Legends

Figure 1 | (A) Mean receiver operating characteristic (ROC) curves derived from the prediction of pCMS using an artificial neural network (ANN), logistic regression (LR), and external models. (B) Mean classification performance parameters achieved using 10x10 stratified cross-validation to predict pCMS on the ANN, LR, and external models. Optimal metrics are highlighted in blue. The ANN performed better than Liu et al.'s model across all metrics ($p<0.001$). It also performed better than Dhaenens et al.'s model in terms of the AUC, accuracy, sensitivity, and negative predictive value ($p<0.001$), though performed worse in terms of specificity and positive predictive value ($p<0.001$). Against logistic regression, the ANN performed better in terms of the AUC and accuracy ($p<0.05$) as well as the sensitivity and negative predictive value ($p<0.001$), though the specificity and positive predictive value did not reach statistical significance. Abbreviations: AUC – area under the curve; PPV – positive predictive value; NPV – negative predictive value.

Figure 2 | Distribution of logistic regression coefficients over cross-validation folds. Arrows indicate whether a risk factor is protective or predictive of pCMS. Coefficients represent the log odds ratio of developing pCMS given a certain risk factor. The log odds of developing pCMS for each tumour type is relative to the modal class (medulloblastoma) whilst the log odds of developing pCMS given compression, signal change, or infiltration of a certain structure is relative to the midline (e.g., compression of the right cerebellar hemisphere is protective in comparison to midline cerebellar compression). Of particular note, the log odds of developing pCMS given involvement of the right / left cerebellar peduncles are relative to involvement of a set of hypothetical midline cerebellar peduncles. In consequence, involvement of the right / left cerebellar peduncles appears falsely protective for pCMS – it is only protective in comparison to this set of (more predictive) hypothetical midline cerebellar peduncles. This is intuitive as midline tumours are more commonly implicated in pCMS, and so they are more likely to affect a set of hypothetical midline cerebellar peduncles. Hence, involvement of the cerebellar peduncles does significantly predispose a child to developing pCMS, with more midline involvement indicative of more severe risk.

Figure 3 | (A-D) Preoperative brain MRI of a 6.5-year-old boy showed a fourth ventricular medulloblastoma with corresponding restricted diffusion (C – apparent diffusion coefficient map). Axial T2-weighted MRI (A) and coronal T2-weighted-fluid-attenuated inversion recovery (B) show compression of the right middle cerebellar peduncle; infiltration of the cerebellar vermis, left cerebellar hemisphere and left cerebellar peduncles; and compression of the right superior cerebellar peduncle. Sagittal T1-weighted post-contrast imaging (D) shows compression of the brainstem and midbrain. Moderate hydrocephalus is also noted (B, D). Given these anatomical and imaging characteristics, the ANN predicted that the patient would develop pCMS (likelihood 90.6%). Clinically, the child subsequently underwent gross total resection via a transvermian approach and manifested symptoms of pCMS in line with this prediction.

(E-H) Preoperative brain MRI of a 4.5-year-old girl showed a large cystic lesion within the posterior fossa with high apparent diffusion coefficient values (G) and an enhancing mural nodule (H – sagittal T1-weighted post-contrast imaging). These imaging features are in keeping with a pilocytic astrocytoma. Axial T2-weighted MRI (E) and coronal T1-weighted inversion recovery (F) show infiltration of the right middle cerebellar peduncle and compression of the brainstem, vermis, fourth ventricle, and left cerebellar hemisphere. Mild hydrocephalus is also noted (F). Given these anatomical and imaging characteristics, the ANN predicted that the patient would not develop pCMS (likelihood 24.1%). Clinically, the child subsequently underwent gross total resection via a trans-cerebellar approach and did not manifest any symptoms of pCMS in line with this prediction.

(I-L) Preoperative brain MRI of an 8.0-year-old girl showed a caudal intraventricular medulloblastoma with compression of the brainstem and cerebellar vermis (I – axial T2-weighted MRI), corresponding restricted diffusion (J – apparent diffusion coefficient map), and moderate hydrocephalus (K – coronal T1-weighted MRI; L – sagittal T1-weighted post-contrast imaging). Given these anatomical and imaging characteristics, the ANN predicted that the patient would not develop pCMS (likelihood 2.4%). Clinically, the child underwent gross total resection via a telovelar approach and did not manifest any symptoms of pCMS in line with this prediction.

		Definition	Interobserver agreement
Qualitative variables			**Fleiss's κ**
Tumour type		Preoperative radiological diagnosis: medulloblastoma, ependymoma, pilocytic astrocytoma, atypical teratoid rhabdoid tumour, other (detail).	0.918
Tumour location		Vermian, caudal / rostral intraventricular, right / left hemispheric, brainstem, other. Tumours may be inputted with more than one location (e.g., large fourth ventricular tumours can be inputted as both caudal and rostral).	0.723
Metastatic at presentation		Presence or absence of brain metastases.	0.825
Preoperative hydrocephalus	Presence	Ventricular enlargement.	0.666
	Grade	Mild – no periventricular signal change or transependymal oedema. Moderate – with transependymal oedema. Severe – compression of external CSF spaces and the cerebral cortex.	0.614
Brainstem	Compression	Distortion of normal brainstem anatomy including anteroposterior displacement against the clivus; effacement of the prepontine / medullary cisterns; and loss of the pontomedullary sulcus.	0.863
	Infiltration	Blurring of the boundary between the brainstem parenchyma and tumour with frank extension of the tumour beyond this boundary.	0.780
Midbrain	Compression	Distortion of normal midbrain anatomy including anterior / superior displacement of the tectal plate.	0.731
	Infiltration	Blurring of the boundary between the midbrain parenchyma and tumour with frank extension of the tumour beyond this boundary.	0.747
Vermis	Compression	Distortion of normal vermian anatomy including effacement of ipsilateral vermian sulci or external cerebrospinal fluid spaces.	0.612
	Infiltration	Tumour may arise from the vermis or there may be lack of distinction between the margin of the vermis and the tumour, with tumoral extension beyond this margin.	0.652
Fourth ventricle	Compression	Effacement by extrinsic tumour or direct infiltration by intrinsic tumour.	0.777
	Infiltration	Presence of tumour within the fourth ventricle or direct invasion of extrinsic tumour to involve the walls of the fourth ventricle.	0.838
Cerebellar hemispheres	Right compression	Distortion of normal cerebellar hemispheric anatomy including effacement of ipsilateral cerebellar sulci or external CSF spaces.	0.623
	Left compression		0.631
	Right infiltration	Tumour may arise from the cerebellar hemispheres or there may be a lack of distinction between the margin of the cerebellar parenchyma and the tumour, with tumoral extension beyond the dentate nuclei and MCPs.	0.760
	Left infiltration		0.755
Middle cerebellar peduncles (MCPs)	Right compression	Distortion of normal MCP anatomy including dorsoventral thinning of the MCP.	0.736
	Left compression		0.761
	Right signal change	Signal change within the MCPs without frank infiltration or blurring of the parenchyma-tumour boundary.	0.683
	Left signal change		0.719
	Right infiltration	Tumour may arise from the MCPs or there may be lack of distinction	0.733

	Left infiltration	between the margin of the MCPs and the tumour, with tumoral extension beyond this margin.	0.715
Superior cerebellar peduncles (SCPs)	Right compression	Distortion of normal SCP anatomy including dorsoventral thinning of the SCP.	0.709
	Left compression		0.687
	Right signal change	Signal change within the SCPs without frank infiltration or blurring of the parenchyma-tumour boundary.	0.675
	Left signal change		0.674
	Right infiltration	Tumour may arise from the SCPs or there may be lack of distinction between the margin of the SCPs and the tumour, with tumoral extension beyond this margin.	0.622
	Left infiltration		0.617
Dentate nuclei *	Right signal change	Signal change within the dentate nuclei without frank infiltration or blurring of the parenchyma-tumour boundary.	0.335
	Left signal change		0.364
Red nuclei *	Right signal change	Signal change within the red nuclei without frank infiltration or blurring of the parenchyma-tumour boundary.	0.493
	Left signal change		0.371
Inferior olivary nuclei *	Right compression	Distortion of normal inferior olivary anatomy including anteroposterior flattening.	0.445
	Left compression		0.309
	Right signal change	Signal change within the inferior olivary nuclei without frank infiltration or blurring of the parenchyma-tumour boundary.	0.373
	Left signal change		0.412
Quantitative variables			**ICC**
Evan's index		A quantitative measure of hydrocephalus severity calculated by dividing the maximal axial diameter of the frontal horns by the maximum intracranial diameter at the same axial level.	0.989
***d*Sag**		Anteroposterior displacement and/or invasion of the brainstem by tumour. If no displacement is present, *d*Sag = 0.	0.997
Maximum cerebellar width		A measure of cerebellar size measured as the greatest transverse axial diameter of the cerebellum.	0.997
Maximum tumour diameter	Anteroposterior	Greatest anteroposterior tumour length as measured on axial MRI.	0.998
	Transverse	Greatest transverse axial width of the tumour.	0.989
	Superoinferior	Greatest superoinferior tumour height perpendicular to the AC-PC line as measured on midline sagittal MRI.	0.999

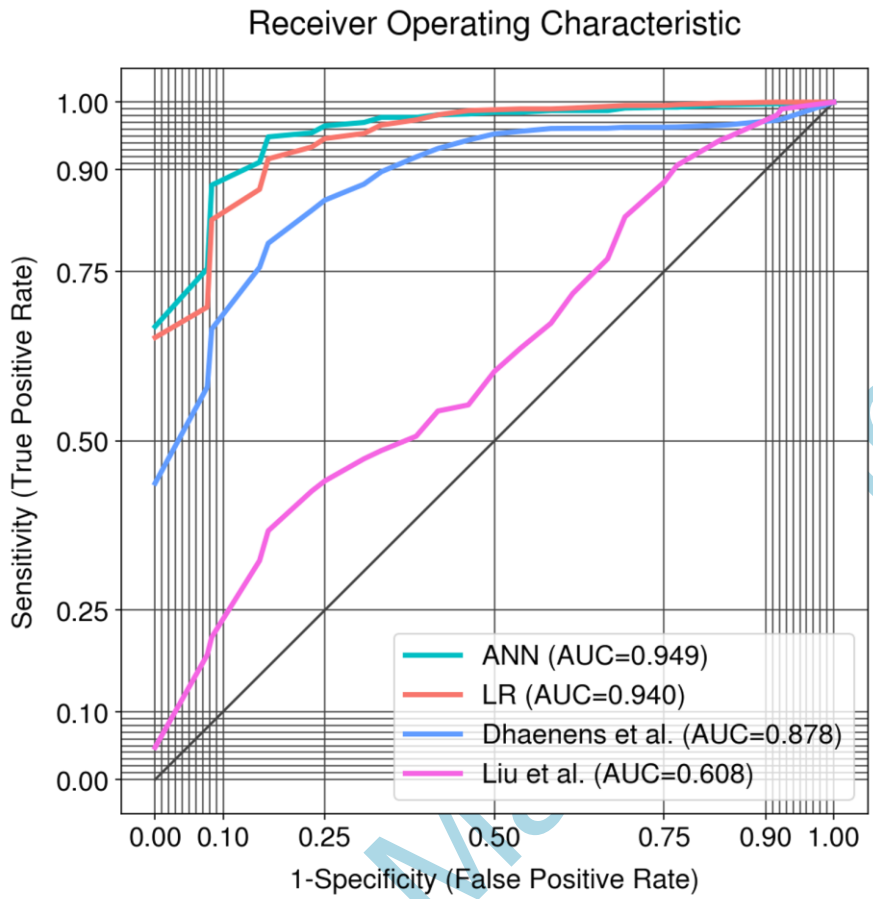
Table 1 | Definition of input variables included in data collection with interobserver agreement determined by Fleiss's κ for qualitative variables and by the intraclass correlation coefficient (ICC) for quantitative variables.[12,66] Inputs denoted by an asterisk (*) were not included in the final model due to low interobserver agreement. Structures infiltrated by tumour are also considered to be compressed by tumour (e.g., fourth ventricular invasion would also be considered as de facto fourth ventricular compression). In cases of potential indistinction, and particularly in larger tumours, we erred towards defining a structure as

infiltrated rather than compressed. Displacement alone without anatomical distortion does not qualify as compression. Signal change may be caused by hydrocephalus, perilesional oedema, or direct interaction with the tumour.[12] For ease of exposition, measurements are shown diagrammatically in Supplementary Figure 2 and exemplar cases of compression, signal change, and infiltration are shown in Supplementary Figure 3. Abbreviations: AC – anterior commissure; CSF – cerebrospinal fluid; MCP – middle cerebellar peduncle; MRI – magnetic resonance imaging; PC – posterior commissure; SCP – superior cerebellar peduncle.

		Cohort			P-value
		All	pCMS	Non-pCMS	
Number of patients enrolled		204	80	124	**< 0.00001**
Mean age (SD) (years)		5.92 (3.88)	5.19 (3.69)	6.17 (4.09)	0.20
Median age (interquartile range) (years)		5.12 (2.61 – 8.64)	4.46 (2.63-7.09)	5.74 (2.52-8.90)	-
Male : female ratio		1.58 : 1	2.20 : 1	1.30 : 1	**0.08**
Tumour type	Medulloblastoma	108	48	60	0.18
	Pilocytic astrocytoma	49	16	33	
	Ependymoma	31	9	22	
	Atypical teratoid rhabdoid tumour	10	5	5	
	Other	6	2	4	
Surgical approach	Transvermian	102	44	58	**0.02**
	Telovelar	69	29	40	
	Other	11	0	11	
	Unknown	22	7	15	
Extent of resection	Gross total resection	146	56	90	0.55
	Subtotal resection	51	22	29	
	Unknown	7	2	5	

Table 2 | Patient demographics with descriptive statistical analysis using t-tests for continuous variables and chi-square tests for categorical variables. Significant values ($p<0.05$) are given in bold.

Figure 1



	Our Models		External Models	
	ANN	LR	Dhaenens et al.	Liu et al.
AUC	0.949 (0.939 – 0.959)	0.940 (0.931 – 0.950)	0.878 (0.862 – 0.894)	0.608 (0.584 – 0.632)
Accuracy	0.909 (0.898 – 0.920)	0.898 (0.886 – 0.910)	0.887 (0.883 – 0.891)	0.398 (0.375 – 0.421)
Sensitivity	0.770 (0.739 – 0.801)	0.693 (0.660 – 0.725)	0.225 (0.197 – 0.253)	0.512 (0.480 – 0.545)
Specificity	0.933 (0.920 – 0.945)	0.933 (0.920 – 0.946)	1.000 (1.000 – 1.000)	0.379 (0.352 – 0.406)
PPV	0.727 (0.683 – 0.770)	0.722 (0.673 – 0.772)	0.960 (0.922 – 0.998)	0.126 (0.118 – 0.134)
NPV	0.960 (0.955 – 0.965)	0.947 (0.941 – 0.953)	0.883 (0.880 – 0.887)	0.802 (0.782 – 0.822)

Figure 2a

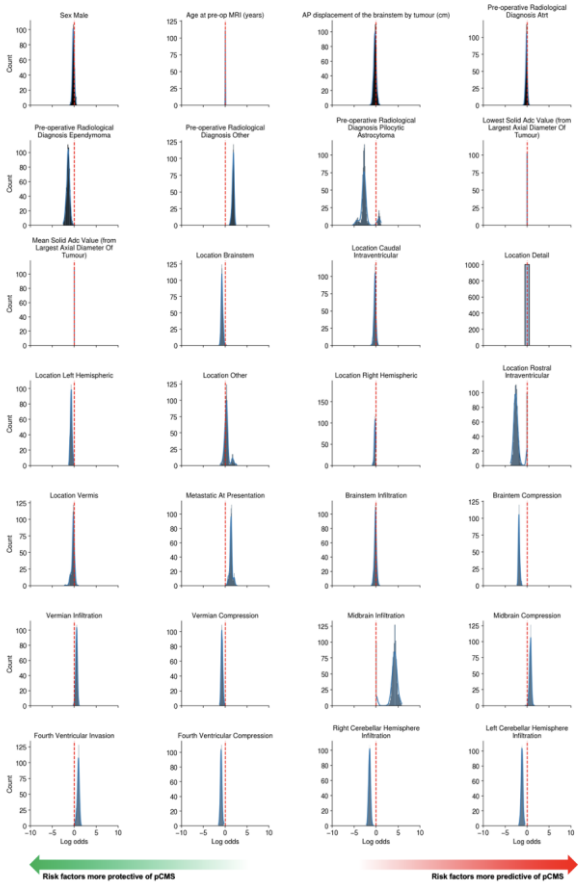
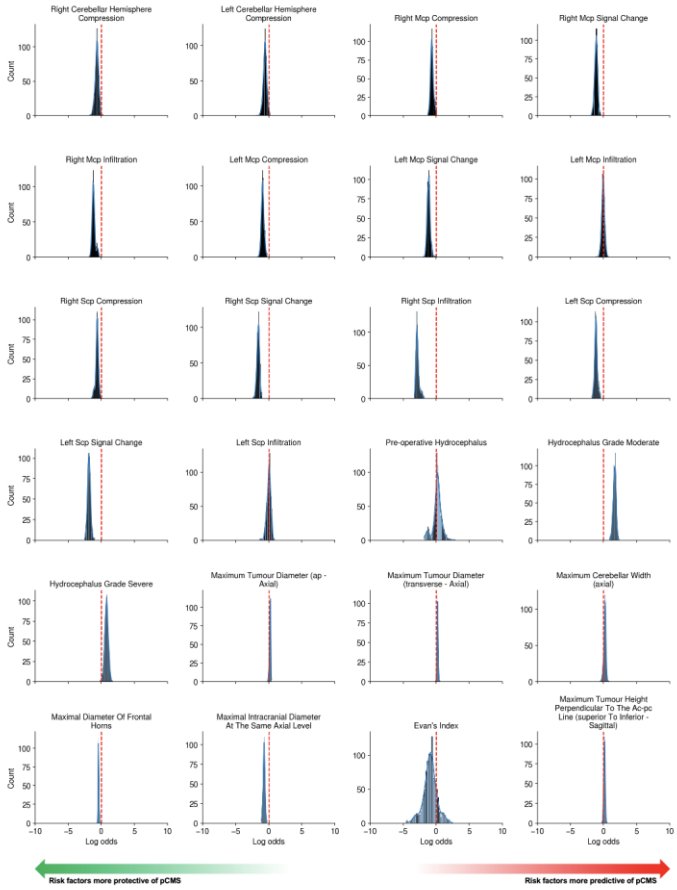


Figure 2b



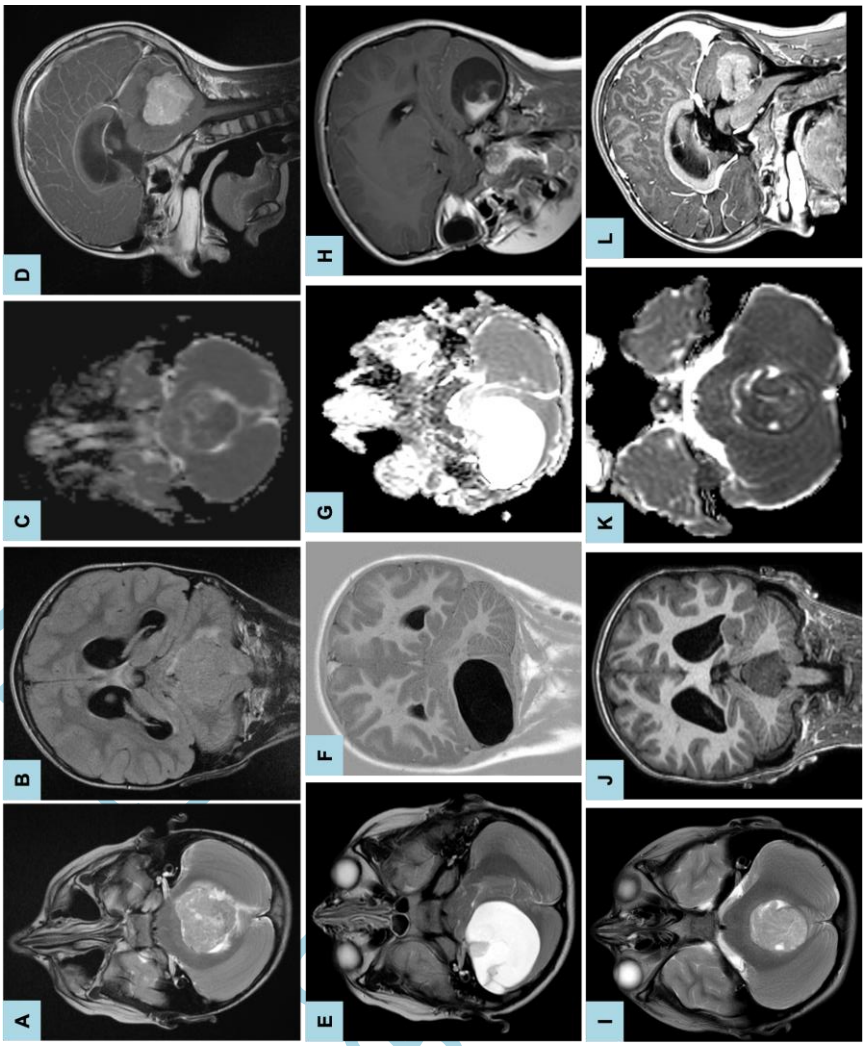


Figure 3