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

Jai Sidpra1,2,3, Adam P. Marcus4, Ulrike Löbel3, Sebastian M. Toescu5,6, Derek Yecies7, Gerald Grant7, Kristen Yeom8, David M. Mirsky9, Hani J. Marcus10,11, Kristian Aquilina2,5, Kshitij Mankad2,3

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.)

© The Author(s) 2022. Published by Oxford University Press, the Society for Neuro-Oncology and the European Association of Neuro-Oncology.
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (https://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com
Department of Neurosurgery, Lucile Packard Children’s Hospital, Stanford, CA 94304, USA (D.Y. and G.G.)

Department of Neuroradiology, Lucile Packard Children’s Hospital, Stanford, CA 94304, USA (K.Y.)

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

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

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


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

Manuscript draft: J.S.

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.\textsuperscript{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.\textsuperscript{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\%.\textsuperscript{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.\textsuperscript{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.\textsuperscript{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.\textsuperscript{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.\textsuperscript{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.\textsuperscript{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.\textsuperscript{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.\textsuperscript{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. Modelling methods and results are reported in line with the Transparent Reporting of A Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) statement.

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. 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 $\kappa$, 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. 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.\textsuperscript{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. 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.  

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. 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. 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%. 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. 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%). Indeed, the comparatively weaker classification performance of decision trees when compared to logistic regression and ANNs has been shown empirically.
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.\textsuperscript{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.\textsuperscript{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.\textsuperscript{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


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

<table>
<thead>
<tr>
<th>Tumour type</th>
<th>All</th>
<th>pCMS</th>
<th>Non-pCMS</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medulloblastoma</td>
<td>108</td>
<td>48</td>
<td>60</td>
<td></td>
</tr>
<tr>
<td>Pilocytic astrocytoma</td>
<td>49</td>
<td>16</td>
<td>33</td>
<td>0.18</td>
</tr>
<tr>
<td>Ependymoma</td>
<td>31</td>
<td>9</td>
<td>22</td>
<td></td>
</tr>
<tr>
<td>Atypical teratoid rhabdoid tumour</td>
<td>10</td>
<td>5</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>6</td>
<td>2</td>
<td>4</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Surgical approach</th>
<th>All</th>
<th>pCMS</th>
<th>Non-pCMS</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transvermian</td>
<td>102</td>
<td>44</td>
<td>58</td>
<td>0.02</td>
</tr>
<tr>
<td>Telovelar</td>
<td>69</td>
<td>29</td>
<td>40</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>11</td>
<td>0</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>Unknown</td>
<td>22</td>
<td>7</td>
<td>15</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Extent of resection</th>
<th>All</th>
<th>pCMS</th>
<th>Non-pCMS</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gross total resection</td>
<td>146</td>
<td>56</td>
<td>90</td>
<td></td>
</tr>
<tr>
<td>Subtotal resection</td>
<td>51</td>
<td>22</td>
<td>29</td>
<td>0.55</td>
</tr>
<tr>
<td>Unknown</td>
<td>7</td>
<td>2</td>
<td>5</td>
<td></td>
</tr>
</tbody>
</table>
Figure 1

Receiver Operating Characteristic

<table>
<thead>
<tr>
<th></th>
<th>Our Models</th>
<th>External Models</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ANN</td>
<td>LR</td>
</tr>
<tr>
<td>AUC</td>
<td>0.949 (0.939 – 0.959)</td>
<td>0.940 (0.931 – 0.956)</td>
</tr>
<tr>
<td>Accuracy</td>
<td>0.909 (0.898 – 0.920)</td>
<td>0.898 (0.886 – 0.910)</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>0.770 (0.739 – 0.801)</td>
<td>0.693 (0.660 – 0.725)</td>
</tr>
<tr>
<td>Specificity</td>
<td>0.933 (0.920 – 0.945)</td>
<td>0.933 (0.920 – 0.946)</td>
</tr>
<tr>
<td>PPV</td>
<td>0.727 (0.683 – 0.770)</td>
<td>0.722 (0.673 – 0.772)</td>
</tr>
<tr>
<td>NPV</td>
<td>0.960 (0.955 – 0.965)</td>
<td>0.947 (0.941 – 0.953)</td>
</tr>
</tbody>
</table>