Child's Nervous System (2021) 37:1485–1494
https://doi.org/10.1007/s00381-021-05061-7

FOCUS SESSION

Machine learning predicts risk of cerebrospinal fluid shunt failure in children: a study from the hydrocephalus clinical research network

Andrew T. Hale[1,2] · Jay Riva-Cambrin[3] · John C. Wellons[2,4] · Eric M. Jackson[5] · John R. W. Kestle[6] · Robert P. Naftel[2,4] · Todd C. Hankinson[7] · Chevis N. Shannon[2,4] · Hydrocephalus Clinical Research Network

Received: 5 January 2021 / Accepted: 22 January 2021 / Published online: 30 January 2021

Abstract

Purpose While conventional statistical approaches have been used to identify risk factors for cerebrospinal fluid (CSF) shunt failure, these methods may not fully capture the complex contribution of clinical, radiologic, surgical, and shunt-specific variables influencing this outcome. Using prospectively collected data from the Hydrocephalus Clinical Research Network (HCRN) patient registry, we applied machine learning (ML) approaches to create a predictive model of CSF shunt failure.

Methods Pediatric patients (age < 19 years) undergoing first-time CSF shunt placement at six HCRN centers were included. CSF shunt failure was defined as a composite outcome including requirement for shunt revision, endoscopic third ventriculostomy, or shunt infection within 5 years of initial surgery. Performance of conventional statistical and 4 ML models were compared.

Results Our cohort consisted of 1036 children undergoing CSF shunt placement, of whom 344 (33.2%) experienced shunt failure. Thirty-eight clinical, radiologic, surgical, and shunt-design variables were included in the ML analyses. Of all ML algorithms tested, the artificial neural network (ANN) had the strongest performance with an area under the receiver operator curve (AUC) of 0.71. The ANN had a specificity of 90% and a sensitivity of 68%, meaning that the ANN can effectively rule-in patients most likely to experience CSF shunt failure (i.e., high specificity) and moderately effective as a tool to rule-out patients at high risk of CSF shunt failure (i.e., moderately sensitive). The ANN was independently validated in 155 patients (prospectively collected, retrospectively analyzed).

Conclusion These data suggest that the ANN, or future iterations thereof, can provide an evidence-based tool to assist in prognostication and patient-counseling immediately after CSF shunt placement.

Keywords Hydrocephalus · CSF shunt failure · HCRN · Machine learning · Artificial intelligence

Introduction

Cerebrospinal fluid (CSF) shunt failure for the treatment of pediatric hydrocephalus is associated with significant morbidity and substantial healthcare costs [1]. The cost of treating hydrocephalus is exceptionally high, especially given the number of readmissions and revision surgeries associated with CSF shunt complications [1]. Thus, identification of

Previously presented as a platform talk at the AANS/CNS Pediatric Section Annual Meeting, Scottsdale AZ, December 8th,2019.

✉ Andrew T. Hale
andrew.hale@vanderbilt.edu

[1] Medical Scientist Training Program, Vanderbilt University School of Medicine, 2200 Pierce Ave., Light Hall 514, Nashville, TN 37232, USA

[2] Surgical Outcomes Center for Kids, Monroe Carell Jr. Children's Hospital of Vanderbilt University, Nashville, TN, USA

[3] Department of Clinical Neurosciences, Alberta Children's Hospital, University of Calgary, Calgary, Alberta, Canada

[4] Division of Pediatric Neurosurgery, Monroe Carell Jr. Children's Hospital of Vanderbilt University, Nashville, TN, USA

[5] Department of Neurosurgery, Johns Hopkins Children's Center, Johns Hopkins University School of Medicine, Baltimore, MD, USA

[6] Department of Neurosurgery, University of Utah, Salt Lake City, UT, USA

[7] Division of Pediatric Neurosurgery, Children's Hospital Colorado, Aurora, CO, USA

patients most likely to experience this outcome may assist in reducing morbidity and costs associated with CSF shunt failure. A number of studies have investigated risk factors for CSF shunt failure; however, critical limitations include retrospective data collection, heterogeneous populations, and small sample size [2, 3]. To overcome these limitations, Riva-Cambrin et al. conducted a prospective cohort study to identify risk factors for CSF shunt failure across six centers within the Hydrocephalus Clinical Research Network (HCRN) [4]. These authors identified age < 6 months at CSF shunt placement, cardiac comorbidity, and endoscopic CSF shunt placement as independently associated with reduced CSF shunt survival using conventional statistical approaches. However, as hydrocephalus is a heterogeneous disorder [5], contributions to CSF shunt failure are multifactorial and may require nuanced approaches to capture small, but clinically meaningful contributions to CSF shunt failure risk. To model this complexity, we compare the performance of conventional statistical approaches and machine learning (ML) algorithms to predict risk of CSF shunt failure. Creation of a predictive model of CSF shunt failure can provide an evidence-based tool for prognostication and assist in counseling patients and their families of their risk of CSF shunt failure at time of surgery.

ML algorithms have been widely used to predict various outcomes in clinical medicine [6] and neurosurgery [7, 8]. Numerous studies have championed ML approaches in neurosurgical topics including pre-operative identification of tumor grade based on features seen on magnetic resonance imaging, prognosticating outcomes after traumatic brain injury, and identifying imaging-based features to predict the need for permanent CSF diversion [9–13]. ML algorithms tend to be more predictive than conventional statistical approaches (i.e., multivariate regression) because they are designed to handle high dimensional inputs and many have demonstrated that ML algorithms can outperform models built using conventional statistical approaches [6–8, 10, 12, 14, 15]. These features enable ML algorithms to more accurately model and predict complex clinical outcomes. However, data-driven selection of the appropriate ML algorithm for the clinical question and unbiased appraisal of the algorithm's strengths/weaknesses is essential. We hypothesized that ML approaches may more accurately model the complexities of CSF shunt failure.

Methods

Patient cohort

Our prospective cohort included children under 19 years of age undergoing initial CSF shunt placement at 6 HCRN centers (Utah/Primary Children's Hospital, University of Alabama at Birmingham/Children's Hospital of Alabama, University of Washington/Seattle Children's Hospital, Washington University/St. Louis Children's Hospital, Baylor College of Medicine/Texas Children's Hospital, and University of Toronto/The Hospital for Sick Children) between April 2008 and December 2011, where centers adhered to a standardized shunt protocol [16]. These data are from the HCRN Core Data Project: Characterizing Patient Populations in the HCRN (Registry). Exclusion criteria included prior placement of CSF shunt or insertion of subdural-peritoneal shunt for subdural hematoma treatment. Patients undergoing subdural-peritoneal shunt placement for subdural hygromas and hydrocephalus were included [4, 17]. Institutional review board (IRB# 131869) approval was obtained from Vanderbilt University (and all other clinical sites) under the HCRN Core Data Project: Characterizing Patient Populations in the HCRN (Registry).

Primary outcome

CSF shunt failure was defined as a composite outcome including requirement for shunt revision, endoscopic third ventriculostomy, or shunt infection within 5 years of initial CSF shunt placement. Diagnosis of shunt infection included one or more of the following: (1) positive identification of bacterial organisms by Gram stain or culture from CSF, pseudocyst, or wound debridement; (2) shunt erosion, defined as the emergence of shunt hardware upon wound complications; (3) abdominal pseudocyst (does not need to include identification of organism); (4) positive blood culture if patient had ventriculoatrial shunt, as previously reported [4]. All variables tested are included in Table 1.

Statistics and machine learning analysis

First, we performed descriptive univariate statistics (Chi-square for categorical variables and Wilcoxon signed-rank tests for continuous variables) to determine any univariate association with CSF shunt failure. Fisher's exact test with Monte Carlo approximation was used for tables greater than 2×2. We tested a number of ML algorithms including coarse Gaussian support vector machines (SVM) [18], naïve Bayesian, k-nearest neighbor algorithms (kNN, 5 neighbors, and 10 neighbors), and artificial neural network (ANN) as previously described [12]. The performance of each model was assessed by the area under the receiver operating characteristic curve (AUC). Construction of the artificial neural network (ANN) was performed as previously described [10–12]. Performance characteristics of the final ANN were tabulated using the confusion matrix [10–12]. We randomly partitioned patients into three groups in order to provide independent validation of our algorithm. The groups were determined as follows: (1) 70% of the cohort was used for model building and training; (2) 15% of the cohort was used for testing of the initial model; (3) 15% of the cohort was used to independently

Table 1 Demographic characteristics of patient cohort. *p* values were calculated using the Wilcoxon rank-sum test[1] or Fisher's exact test[2] with Monte Carlo approximation for tables larger than 2 × 2. Data are presented as mean ± SEM. Abbreviations are as follows: intraventricular hemorrhage (*IVH*); intracerebral hemorrhage (*ICH*); subarachnoid hemorrhage (*SAH*); central nervous system (*CNS*); endoscopic third ventriculostomy (*ETV*); entraventricular drain (*EVD*); frontal occipital horn ratio (*FOHR*). Hydrocephalus due to tumors include posterior fossa, supratentorial, and midbrain lesions. Congenital causes of hydrocephalus include congenital communicating, encephalocele, and craniosynostosis

	No CSF shunt failure (*n*= 692)	CSF shunt failure (*n*= 344)	*p* value*
Demographic			
Age (weeks)	24.0 (7.0.–188.5)	14.0 (2.0–69.0)	< 0.001[1]
Sex			0.126[2]
Male	377 (54.5%)	205 (59.6%)	
Female	315 (45.5%)	139 (40.4%)	
Birth weight (kg)	3.0 (1.74–3.40)	2.90 (1.24–3.35)	0.108[1]
Weight at surgery (kg)	6.90 (3.55–15.70)	4.30 (3.00–10.40)	< 0.001[1]
Gestational age at birth	38.0 (32.0–40.0)	37.0 (28.0–39.0)	0.023[1]
Race			0.616[2]
Caucasian	446 (64.5%)	209 (60.8%)	
African-American	102 (14.7%)	50 (14.5%)	
Asian	14 (2.0%)	6 (1.7%)	
American Indian	6 (0.9%)	1 (0.3%)	
Pacific islander	2 (0.3%)	1 (0.3%)	
Other	12 (1.7%)	7 (2.0%)	
Unknown	110 (15.9%)	70 (20.3%)	
Ethnicity			0.776[2]
Non-Hispanic	597 (86.3%)	294 (85.5%)	
Hispanic	95 (13.7%)	50 (14.5%)	
Insurance			0.320[2]
Public (Medicare, Medicaid)	372 (53.8%)	198 (57.6%)	
Private	284 (41.0%)	125 (36.3%)	
Other (i.e., military)	36 (5.2%)	21 (6.1%)	
Season			0.425[2]
January–March	138 (19.9%)	60 (17.4%)	
April–June	153 (22.1%)	91 (26.5%)	
July–September	201 (29.0%)	94 (27.3%)	
October–December	200 (28.9%)	99 (28.8%)	
Clinical variables			
Hydrocephalus etiology			0.004[2]
Post-infectious	25 (3.6%)	13 (3.8%)	
IVH of prematurity	135 (19.5%)	91 (26.5%)	
Myelomeningocele	93 (13.4%)	71 (20.6%)	
Aqueductal stenosis	60 (8.7%)	24 (7.0%)	
Spontaneous IVH, ICH, or SAH	30 (4.3%)	9 (2.6%)	
Tumor	135 (19.5%)	57 (16.6%)	
Post-head injury	35 (5.1%)	14 (4.1%)	
Intracranial cyst	56 (8.1%)	28 (8.1%)	
Congenital	95 (13.7%)	31 (9.0%)	
Other	28 (4.0%)	6 (1.7%)	
Gastrostomy tube			0.315[2]
Yes	62 (9.0%)	38 (11.0%)	
No	630 (91.0%)	306 (89.0%)	
Endotracheal tube			0.229[2]
Yes	16 (2.3%)	13 (3.8%)	
No	676 (97.7%)	331 (96.2%)	
Neuromuscular condition			0.403[2]
Yes	81 (11.7%)	34 (9.9%)	
No	611 (88.3%)	310 (90.1%)	
Cardiovascular comorbidity			0.088[2]
Yes	48 (6.9%)	35 (10.2%)	
No	644 (93.1%)	89.8%)	
Gastrointestinal comorbidity			0.061[2]
Yes	7 (1.0%)	9 (2.6%)	
No	685 (93.1%)	335 (97.4%)	
Renal comorbidity			0.490[2]
Yes	5 (0.7%)	4 (1.2%)	
No	687 (99.3%)	340 (98.8%)	
Known congenital or genetic defect			0.758[2]
Yes	35 (5.1%)	15 (4.4%)	
No	657 (94.9%)	329 (95.6%)	

Table 1 (continued)

	No CSF shunt failure (*n*= 692)	CSF shunt failure (*n*= 344)	*p* value*
History of non-CNS malignancy			0.301[2]
Yes	21 (3.0%)	6 (1.7%)	
No	671 (97.0%)	338 (98.3%)	
Length of stay (days)	2.0 (0.0–13.0)	4.0 (1.0–18.0)	< 0.001[1]
Surgical and radiologic variables			
Surgeon volume (shunts per year)	17.5 (12.9–20.05)	17.4 (14.6–20.02)	0.935[1]
Endoscopically assisted			< 0.001[2]
Yes	150 (21.7%)	121 (35.2%)	
No	542 (78.3)	235 (68.3%)	
Intraoperative ultrasound			0.778[2]
Yes	226 (32.7%)	109 (31.7%)	
No	466 (67.3%)	235 (68.3%)	
Frameless stereotaxy			0.728[2]
Yes	35 (5.1%)	14 (4.1%)	
No	212 (30.6%)	75 (21.8%)	
Unknown	445 (64.3%)	255 (74.1%)	
Case duration (minutes)	51.0 (40.0–66.0)	52.0 (39.0–68.0)	0.809[1]
Surgical priority			0.231[2]
Elective	435 (62.9%)	218 (63.4%)	
Add-on	229 (33.1%)	119 (34.6%)	
Emergent	28 (4.0%)	7 (2.0%)	
ETV concurrent with shunt			0.691[2]
Yes	4 (0.6%)	3 (0.9%)	
No	688 (99.4%)	341 (99.1%)	
History of ETV			< 0.001[2]
Yes	1 (0.1%)	8 (2.3%)	
No	691 (99.9%)	336 (97.7%)	
History of subgaleal shunt			0.539[2]
Yes	15 (2.2%)	13 (3.8%)	
No	120 (17.3%)	78 (22.7%)	
Unknown	557 (80.5%)	253 (73.5%)	
History of EVD placement			0.610[2]
Yes	28 (4.0%)	16 (4.7%)	
No	107 (15.5%)	75 (21.8%)	
Unknown	557 (80.5%)	253 (73.5%)	
Days after birth to first ventricular access	32.0 (21.0–46.0)	32.5 (22.5–44.0)	0.918[1]
Prior neurosurgical procedure			0.552[2]
Yes	91 (13.2%)	40 (11.6%)	
No	601 (86.8%)	304 (88.4%)	
FOHR at surgery	0.54 (0.44–0.64)	0.55 (0.45–0.64)	0.435[1]
Surgical and radiologic variables			
Shunt manufacturer			0.020[2]
1	292 (42.2%)	141 (41.0%)	
2	147 (21.2)	109 (31.7%)	
3	77 (11.1%)	30 (8.7%)	
4	51 (7.4%)	19 (5.5%)	
5	50 (7.2%)	22 (6.4%)	
6	31 (4.5%)	5 (1.5%)	
7	18 (2.6%)	10 (2.9%)	
8	4 (0.6%)	1 (0.3%)	
9	4 (0.6%)	1 (0.3%)	
10	2 (0.3%)	2 (0.6%)	
11	1 (0.1%)	0 (0%)	
Missing	4 (0.6%)	0 (0%)	
Shunt type			0.138[2]
Fixed	553 (79.9%)	292 (84.9%)	
Programmable	128 (18.5%)	49 (14.2%)	
Other	11 (1.6%)	3 (0.9%)	
Bactiseal			0.408[2]
Yes	83 (12.0%)	35 (10.2%)	
No	609 (88.0%)	309 (89.8%)	
Proximal catheter location			0.587[2]
Ventricular	645 (93.2%)	323 (93.9%)	
Subdural	19 (2.7%)	6 (1.7%)	
Cyst	17 (2.5%)	7 (2.0%)	

Table 1 (continued)

	No CSF shunt failure (n= 692)	CSF shunt failure (n= 344)	p value*
Complex	5 (0.7%)	6 (1.7%)	
Lumbar	5 (0.7%)	2 (0.6%)	
Peritoneal entry method			0.603[2]
Minilap	396 (57.2%)	193 (56.1%)	
Trocar	238 (34.4%)	124 (36.0%)	
Laproscopic	15 (2.2%)	9 (2.6%)	
Unknown	43 (6.2%)	18 (5.3%)	
Anti-siphon			0.043[2]
Yes	219 (31.6%)	131 (38.1%)	
No	473 (68.4%)	213 (61.9%)	

validate the performance of the ANN as previously described [11, 12]. With this ANN training paradigm established, we iteratively designed one, two, and three hidden-layer, feed-forward neural networks with between zero and fifty artificial neurons per layer. Each of these models was built using the scaled conjugate gradient back propagation algorithm on the dedicated partition, as is standard in the ML literature [19]. This approach identified the most predictive architecture, optimizing the AUC, as an ANN with fifteen, thirty, and twenty artificial neurons, respectively. We tabulated confusion matrix tables and statistics on the validation cohort as previously described [10–12].

Results

Demographic, clinical, surgical, radiologic, and cerebrospinal fluid shunt variables

A total of 1036 patients undergoing initial CSF shunt placement were included in our study. Of those patients, 344 (33.2%) experienced CSF shunt failure, defined as a composite outcome including requirement for shunt revision, endoscopic third ventriculostomy, or shunt infection within 5 years of initial CSF shunt surgery. Overall, younger patients (14.0 ± 2.0–69.0 weeks, corrected gestational age; mean ± standard deviation) were more likely to experience CSF shunt failure than older patients at surgery (24.0 ± 7.0–188.5 weeks, corrected gestational age; mean ± standard deviation, p< 0.001, Table 1). Similarly, patients born at younger corrected gestational age (37.0 ± 28.0–39.0 vs. 37.0 ± 28.0–39.0 weeks, mean ± standard deviation; p= 0.023) and smaller weight at surgery (4.30 ± 3.00–10.40 vs. 6.90 ± 3.55–15.70 kg; mean ± standard deviation, p< 0.001) were more likely to experience CSF shunt failure (Table 1). Sex, birth weight, race, ethnicity, insurance type, and season at time of CSF shunt placement were not associated with CSF shunt failure by univariate analysis.

Next, we considered clinical factors and their association with CSF shunt failure. We identified etiology of hydrocephalus (p= 0.004) and longer length of stay (4.0 ± 1.0–18.0 vs. 2.0 ± 0.0–13.0; mean ± standard deviation, p< 0.001, Table 2) as significantly associated with CSF shunt failure. However, none of the following variables reached statistical significance for their association with CSF shunt failure by univariate analysis: (1) pre-operative placement of a gastrostomy tube, (2) pre-operative placement of endotracheal tube, (3) neuromuscular disease comorbidity, (4) cardiovascular disease comorbidity, (5) gastrointestinal disorder comorbidity, (6) renal comorbidity, (7) presence of known congenital or genetic disease, and (8) history of a non-central nervous system malignancy.

We then considered surgical variables, key events in the surgical history, and radiologic factors that may influence CSF shunt failure. We found that endoscopically assisted CSF shunt placement (p< 0.001) and history of previous endoscopic third ventriculostomy (ETV, p< 0.001, Table 3) were associated with CSF shunt failure by univariate analysis. However, the following variables were not associated with CSF shunt failure (p> 0.05,

Table 2 Performance parameters of machine learning and statistical models tested. Logistic regression was performed using variables reaching p < 0.05 in Table 1 as covariates. Abbreviations are as follows: *kNN*, k-nearest neighbors' algorithm, where k = number of neighbors tested (5 or 10); *SVM*, support vector machine

	AUC
Model	
Logistic regression	0.613
kNN	
5 neighbors	0.622
10 neighbors	0.647
SVM (course gaussian)	0.689
Naïve Bayes	0.670

Table 3 Performance characteristics of artificial neural network (ANN). The ANN was trained using 70% of our cohort (n= 725), algorithm testing and refinement using an additional 15% of patients (n= 156), and 15% of patients (n= 155) were used for validation of the final algorithm. Abbreviations are as follows: positive predictive value (*PPV*), negative predictive value (*NPV*), and area under the receiver operator characteristic curve (*AUC*)

Performance parameter	Testing cohort	Final algorithm
Specificity	85%	90%
Sensitivity	65%	68%
PPV	69%	74%
NPV	83%	87%
AUC	0.68	0.71

Table 3): (1) surgeon volume (defined as number of CSF shunts placed per year), (2) use of intraoperative ultrasound during CSF shunt placement, (3) use of frameless stereotaxy during CSF shunt placement, (4) case duration (minutes), (5) surgical priority (elective, add-on, or emergent), (6) ETV concurrent with CSF shunt placement, (7) history of subgaleal shunt placement, (8) history of extraventricular drain (EVD) placement, (9) days after birth to first ventricular access, (10) history of any prior neurosurgical procedure, and (11) ventricular size as assessed by the frontal occipital horn ratio (FOHR) [20]. We then considered CSF shunt-specific variables and their association with CSF shunt failure (Table 1). By univariate analysis, we identified shunt manufacturer (p= 0.020) and use of an anti-siphon device (38.1 vs. 31.6%) as being associated with CSF shunt failure (Table 1). However, shunt type (fixed, programmable, or unknown), use of an antibiotic-impregnated catheter, proximal catheter location (ventricular, subdural, cyst, complex, or lumbar), and peritoneal entry method (minilap, trocar, laproscopic, or unknown) were not statistically significant predictors of CSF shunt failure.

Machine learning approaches to predict CSF shunt failure

We next aimed to create a predictive model of CSF shunt failure using a variety of approaches. First, we included all statistically significant ($p < 0.05$) variables by univariate analysis (Table 1) in a multivariate logistic regression model (Table 2, AUC = 0.613). Next, we compared a variety of ML approaches including non-parametric approaches k-nearest neighbors (kNN, including 5 or 10 neighbors, AUC = 0.622 and 0.647, respectively) and parametric approaches including coarse Gaussian support vector machine (SVM, AUC = 0.689) and Naïve Bayesian models (AUC = 0.670). Although these ML approaches displayed higher predictive potential than a logistic regression model (Table 2), ANN have also been shown to accurately predict neurosurgical outcomes as well [21], and may more accurately reflect the higher-order complexity of CSF shunt failure and small, yet significant, contribution of many variables. Thus, we decided to construct an ANN and compare its performance to the aforementioned ML algorithms.

Artificial neural network to predict CSF shunt failure

Of all ML approaches tested, the ANN had the highest AUC (0.71, Tables 2 and 3, Fig. 1a). A schematic of final algorithm is shown in Fig. 1b. A detailed description of the final algorithm and iterative approach taken to arrive at that model can be found in the "Methods" section. The ANN was trained using 70% of our cohort (n= 725) and refined in 15% (n= 156) of our cohort. Importantly, the ANN was independently validated on 155 patients (performance characteristics listed in Table 3) before the final algorithm was constructed. Performance characteristics of the final ANN are listed in Table 3. The specificity of the ANN was 90% indicating that the algorithm is effective at identifying patients most likely to experience (i.e., rule in) CSF shunt failure. However, the sensitivity of the ANN was 68% indicating that the performance of the algorithm is only modestly efficacious at ruling-out those likely to experience CSF shunt failure. The positive predictive value (PPV) of the algorithm was 74% and the negative predictive value (NPV) was 87%. The final algorithm displayed a misclassification rate of 17%.

Discussion

Using prospectively collected data from six HCRN centers across North America, we compare performance of conventional statistical and ML models. Of all algorithms built and tested, the ANN performs the best to predict CSF shunt failure in children undergoing initial CSF shunt placement, consistent with prior ML studies in the neurosurgical literature [11, 12]. However, the most important conclusion that can be drawn from these data is the relatively modest performance of the algorithm. Despite testing multiple ML algorithms, using a robust sample size, and including 38 prospectively collected demographic, clinical, radiologic, and surgical variables, the model was 90% specific and 68% sensitive. These data highlight the complex nature of CSF shunt failure risk and our elementary understanding of mitigating factors to prevent device failure. We hypothesize that incorporation of other clinical data elements and additional types of data (i.e., genetic, radiologic) and/or more sophisticated ML

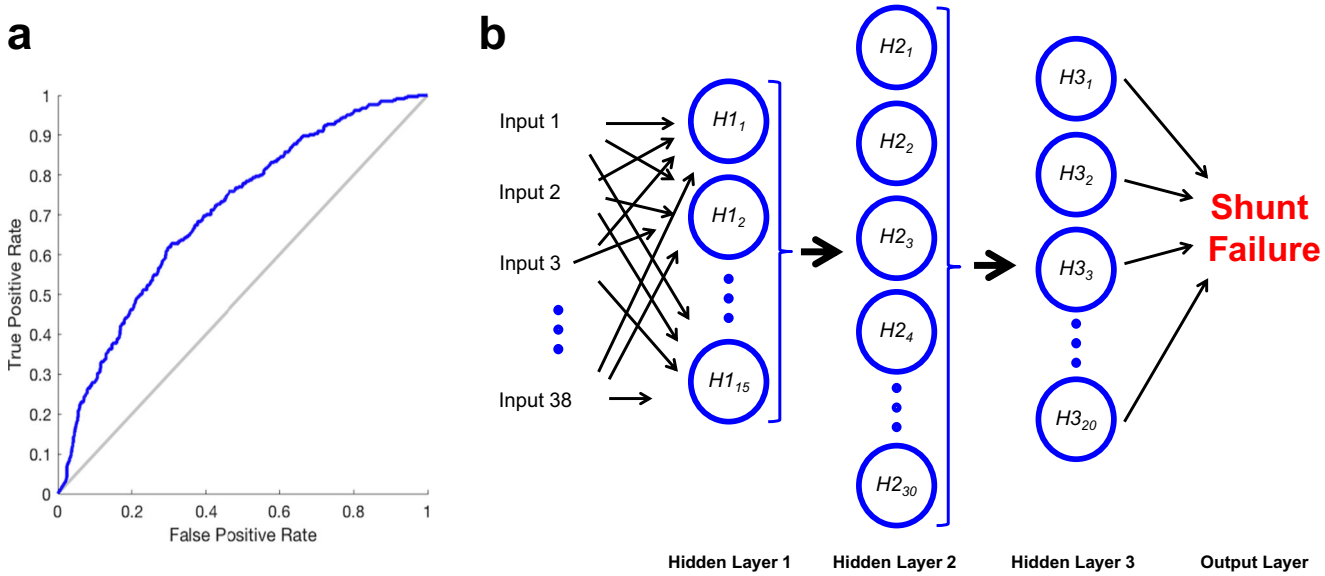

Fig. 1 Performance and representation of the artificial neural network (ANN) to predict CSF shunt failure. **a** Area under the receiver operator curve (AUC) for the final ANN designed to predict CSF shunt failure. **b** Schematic of the ANN constructed here. Thirty-eight input variables were included (listed in Table 1), converging on 15 (first layer) then 30 (second layer) then 20 (third layer) training nodes (for simplicity, fewer nodes are shown). Each input variable projects to each training node. Arbitrary "weights" are then assigned to each variable. Each training node is then used to determine the optimal weights of each variable onto the next layer in order to provide the most robust prediction of CSF shunt failure ("output layer") determined by AUC

methods may increase the performance of future iterations of this algorithm and shed light on potential mechanisms underlying the high failure rate of CSF shunts.

Young age (< 2 years at the time of CSF shunt placement), use of the endoscope, and cardiac comorbidity at the time of CSF shunt placement are independently associated with CSF shunt failure by Cox regression modeling [4]. While conventional statistical approaches are very relevant in identifying potentially modifiable risk factors, these stringent approaches may not fully capture the complex contribution of multiple variables with small effect-size contributing to CSF shunt failure. Compared to conventional statistical approaches, ML algorithms have the advantage of being able to accommodate different types (i.e., continuous and categorical) and an exceedingly large number of variables, enabling a more unbiased approach capturing the complexity of individual patient information [7, 8, 10]. The advantages of ML algorithms over conventional statistical methods have been well documented in the medical and neurosurgical literature [6–8, 10–12]. However, careful selection of the appropriate ML algorithm to apply to a dataset is an important consideration.

Here, we test a number of ML algorithms including Gaussian support vector machines, Naïve Bayesian, k-nearest neighbor algorithms, and ANN to determine the most predictive model. Importantly, we provide validation of our model using data within the HCRN (Table 3 and Fig. 1a–b), which we and others have demonstrated is broadly representative of other centers across North America [22–27]. ANN are constructs best equipped to rationalize large amounts of data by randomly assigning "weights" to different variables within the model. These weights are then compared against all other variables in the model and chosen based on unbiased prediction and correlation with the outcome of interest. Although ANN can handle very large numbers of variables, many have shown that the data included must still be relevant to the outcome of interest [28]. Thus, inclusion of additional data elements heretofore not yet identified that play a role in hydrocephalus and CSF shunt failure risk as well as refinement of ML approaches will improve the algorithm's performance.

We posit that next-generation ML models should focus on three key elements: (1) prospective collection of data and real-time updates of the algorithm based on the most recent evidence; (2) incorporation of ML models in the electronic health record (EHR); and (3) incorporation of additional data elements into the model such as radiologic and genetic information. While conventional statistical models require de novo analysis with incorporation of new data, ML algorithms can be tweaked and refined in real time. In addition, providers may choose to build their own models based on specific factors in their practice, hospital system, or region. Many of the data elements used here could reasonably be extracted from the EHR, limiting the burden on the provider. Finally, additional granular data elements such as direct analysis of radiographic studies and/or incorporation of genetic information may be to increase the performance of the algorithm.

The pathophysiology of hydrocephalus is exceedingly complex [5], and numerous genetic factors have been thought to contribute to hydrocephalus risk [29]. There is an increasing appreciation that hydrocephalus is influenced, at least in part, by genetic factors both alone and in combination with environmental factors [29]. However, the extent to which genetic factors may play a role in CSF shunt failure is unknown. As our understanding of the genetic basis of hydrocephalus increases, it is worth considering whether or not genetic factors play a role in CSF shunt failure. As the pathophysiology of hydrocephalus is also not well understood [5, 30] and there are many etiologies of hydrocephalus which may even share common pathobiological mechanisms [30, 31], discerning the role of genetics is likely to be very fruitful. For instance, it is possible to imagine genetic testing playing a role in hydrocephalus disease management [32], as genetic testing is routinely performed for other neurological diseases including cancer, epilepsy, cerebrovascular disease, and neurodevelopmental disorders [33–37]. Understanding the genetic architecture of both hydrocephalus and CSF shunt failure will undoubtedly increase our understanding of hydrocephalus pathogenesis and will be useful to improve the performance of our ANN and inform creation of hydrocephalus etiology-specific models.

While we provide initial data suggesting that ML algorithms can predict risk of CSF shunt failure, our study is not without limitations. The ANN, along with other ML algorithms, are black box models and cannot be interrogated fully. This is an active area of investigation in ML. However, the large volumes of data that are increasingly becoming available for the study of biomedical challenges mandate analysis methods that are more sophisticated than conventional statistics [38]. While we agree that ML algorithms are conceptually difficult to understand compared to conventional statistical approaches, the emerging importance of "big data" in neurosurgery will undoubtedly necessitate basic understanding of ML algorithms [39, 40]. However, similarly nuanced approaches for "small-data" contexts are also needed to better understand less common diseases [41]. Second, while this represents the largest study of CSF shunt failure, the number of patients with each etiology of hydrocephalus is relatively low. With larger amounts of data, etiology-specific ML algorithms for CSF shunt failure may be constructed further increasing the predictive performance of the model. In addition, we did not consider other factors relevant to permanent CSF diversion, including the option for ETV with choroid plexus cauterization (ETV/CPC), which is also associated with high failure rates [42–46]. However, the decision to pursue CSF shunt placement vs. ETV/CPC will be best answered by ongoing randomized clinical trials (RCT) in North America, Uganda, and elsewhere across the world. Thus, our ANN may only be relevant after the decision to pursue CSF shunt failure is made, but incorporating factors relevant to the decision to pursue CSF shunt placement vs. ETV/CPC may be useful in the future.

Conclusions

Using prospectively collected data from the HCRN, we create the first ML algorithm (ANN) to predict CSF shunt failure in children with hydrocephalus. We hypothesize that incorporation of additional sources of data (i.e., radiologic and/or genetic) may further increase the performance of our model as well as our understanding of CSF shunt failure and, 1 day when acceptably accurate, contribute to point-of-care decision making. The data reported herein provide a framework for future studies using ML algorithms to predict outcomes in pediatric hydrocephalus treatment and further improve the care of the patients we have committed to care for.

Acknowledgements The authors would like to thank their colleagues for their past and ongoing support of HCRN:

D Brockmeyer, M Walker, R Bollo, S Cheshier, J Blount, J Johnston, B Rocque, L Acakpo-Satchivi, WJ Oakes, J Rutka, M Taylor, P Dirks, D Curry, G Aldave, R Dauser, A Jea, S Lam, H Weiner, T Luerssen, R Ellenbogen, J Ojemann, A Lee, A Avellino, S Greene, E Tyler-Kabara, T Abel, TS Park, J Strahle, S McEvoy, M Smyth, N Tulipan, A Singhal, P Steinbok, D Cochrane, W Hader, C Gallagher, M Benour, E Kiehna, JG McComb, P Chiarelli, A Robison, A Alexander, M Handler, B O'Neill, C Wilkinson, L Governale, A Drapeau, J Leonard, E Sribnick, A Shaikhouni, E Ahn, A Cohen, M Groves, S Robinson, and CM Bonfield.

In addition, our work would not be possible without the outstanding support of the dedicated personnel at each clinical site and the data coordinating center. Special thanks go to the following: L Holman, J Clawson, P Martello, N Tattersall, T Bach (Salt Lake City); A Arynchyna, A Bey (Birmingham); H Ashrafpour, M Lamberti-Pasculli, L O'Connor (Toronto); E Sanchez, E Santisbon, S Martinez, S Ryan (Houston); A Anderson, G Bowen (Seattle); K Diamond, A Luther (Pittsburgh); A Morgan, H Botteron, D Morales, M Gabir, D Berger, D Mercer (St. Louis); A Wiseman, J Stoll, D Dawson, S Gannon (Nashville); A Cheong, R Hengel (British Columbia); R Rashid, S Ahmed (Calgary); M Alrefaie, R Daniel, A Loudermilk (Baltimore); N Rea, C Cook (Los Angeles); S Staulcup (Colorado); A Sheline (Columbus); and N Nunn, M Langley, V Wall, D Austin, B Conley, V Freimann, L Herrera, B Miller (Utah Data Coordinating Center).

Funding The HCRN has been funded by National Institute of Neurological Disorders and Stroke (NINDS grant nos. 1U01NS107486-01A1 and 1RC1NS068943-01), Patient Centered Outcome Research Institute (PCORI grant no. CER-1403-13857), The Gerber Foundation (reference no. 1692-3638), private philanthropy, and the Hydrocephalus Association. A.T.H. is supported by the National Institutes of Health (F30HL143826) and Vanderbilt University Medical Scientist Training Program (5T32GM007347). None of the sponsors participated in design and conduct of this study; collection, management, analysis, and interpretation of the data; or preparation, review, or approval of this manuscript. Its contents are solely the responsibility of the authors and do not necessarily represent the official view of the sponsors.

Declarations

Conflict of interest None.

Appendix. Hydrocephalus Clinical Research Network

Members

The HCRN currently consists of the following clinical centers and investigators: Primary Children's Hospital, University of Utah (J Kestle); Children's Hospital of Alabama, University of Alabama at Birmingham (C Rozzelle); Hospital for Sick Children, University of Toronto (J Drake, A Kulkarni); Texas Children's Hospital, Baylor College of Medicine (W Whitehead); Seattle Children's Hospital, University of Washington (S Browd, T Simon, J Hauptman); Children's Hospital of Pittsburgh, University of Pittsburgh (I Pollack); St. Louis Children's Hospital, Washington University in St. Louis (D Limbrick); Monroe Carell Jr. Children's Hospital at Vanderbilt, Vanderbilt University Medical Center (J Wellons, R Naftel, C Shannon); British Columbia Children's Hospital, University of British Columbia (M Tamber, P McDonald); Alberta Children's Hospital, University of Calgary (J Riva-Cambrin); The Johns Hopkins Hospital (E Jackson); Children's Hospital of Los Angeles (M Krieger, J Chu); Children's Hospital Colorado (T Hankinson); Nationwide Children's Hospital (J Pindrik); HCRN Data Coordinating Center, Department of Pediatrics, University of Utah (R Holubkov).

References

1. Lim J, Tang AR, Liles C, Hysong AA, Hale AT, Bonfield CM, Naftel RP, Wellons JC, Shannon CN (2018) The cost of hydrocephalus: a cost-effectiveness model for evaluating surgical techniques 1
2. Lazareff JA, Peacock W, Holly L, Ver Halen J, Wong A, Olmstead C (1998) Multiple shunt failures: an analysis of relevant factors. Childs Nerv Syst 14:271–275
3. Tuli S, Drake J, Lawless J, Wigg M, Lamberti-Pasculli M (2000) Risk factors for repeated cerebrospinal shunt failures in pediatric patients with hydrocephalus. J Neurosurg 92:31–38
4. Riva-Cambrin J, Kestle JR, Holubkov R, Butler J, Kulkarni AV, Drake J, Whitehead WE, Wellons JC 3rd, Shannon CN, Tamber MS, Limbrick DD Jr, Rozzelle C, Browd SR, Simon TD (2016) Risk factors for shunt malfunction in pediatric hydrocephalus: a multicenter prospective cohort study. J Neurosurg Pediatr 17:382–390
5. Tomycz LD, Hale AT, George TM (2017) Emerging insights and new perspectives on the nature of hydrocephalus. Pediatr Neurosurg 52:361–368
6. Rajkomar A, Dean J, Kohane I (2019) Machine learning in medicine. N Engl J Med 380:1347–1358
7. Buchlak QD, Esmaili N, Leveque JC, Farrokhi F, Bennett C, Piccardi M, Sethi RK (2019) Machine learning applications to clinical decision support in neurosurgery: an artificial intelligence augmented systematic review. Neurosurg Rev
8. Senders JT, Staples PC, Karhade AV, Zaki MM, Gormley WB, Broekman MLD, Smith TR, Arnaout O (2018) Machine learning and neurosurgical outcome prediction: a systematic review. World Neurosurg 109:476–486.e471
9. Cherukuri V, Ssenyonga P, Warf BC, Kulkarni AV, Monga V, Schiff SJ (2018) Learning based segmentation of CT brain images: application to postoperative hydrocephalic scans. IEEE Trans Biomed Eng 65:1871–1884
10. Hale AT, Stonko DP, Brown A, Lim J, Voce DJ, Gannon SR, Le TM, Shannon CN (2018) Machine-learning analysis outperforms conventional statistical models and CT classification systems in predicting 6-month outcomes in pediatric patients sustaining traumatic brain injury. Neurosurg Focus 45:E2
11. Hale AT, Stonko DP, Lim J, Guillamondegui OD, Shannon CN, Patel MB (2018) Using an artificial neural network to predict traumatic brain injury. J Neurosurg Pediatr:1–8
12. Hale AT, Stonko DP, Wang L, Strother MK, Chambless LB (2018) Machine learning analyses can differentiate meningioma grade by features on magnetic resonance imaging. Neurosurg Focus 45:E4
13. Pisapia JM, Akbari H, Rozycki M, Goldstein H, Bakas S, Rathore S, Moldenhauer JS, Storm PB, Zarnow DM, Anderson RCE, Heuer GG, Davatzikos C (2018) Use of fetal magnetic resonance image analysis and machine learning to predict the need for postnatal cerebrospinal fluid diversion in fetal ventriculomegaly. JAMA Pediatr 172:128–135
14. Darcy AM, Louie AK, Roberts LW (2016) Machine learning and the profession of medicine. JAMA 315:551–552
15. Obermeyer Z, Emanuel EJ (2016) Predicting the future - big data, machine learning, and clinical medicine. N Engl J Med 375:1216–1219
16. Kestle JR, Riva-Cambrin J, Wellons JC 3rd, Kulkarni AV, Whitehead WE, Walker ML, Oakes WJ, Drake JM, Luerssen TG, Simon TD, Holubkov R (2011) A standardized protocol to reduce cerebrospinal fluid shunt infection: the Hydrocephalus Clinical Research Network Quality Improvement Initiative. J Neurosurg Pediatr 8:22–29
17. Simon TD, Butler J, Whitlock KB, Browd SR, Holubkov R, Kestle JR, Kulkarni AV, Langley M, Limbrick DD Jr, Mayer-Hamblett N, Tamber M, Wellons JC 3rd, Whitehead WE, Riva-Cambrin J (2014) Risk factors for first cerebrospinal fluid shunt infection: findings from a multi-center prospective cohort study. J Pediatr 164:1462–1468.e1462
18. Gestel TV, Suykens JAK, Lanckriet G, Lambrechts A, Moor BD, Vandewalle J (2002) Bayesian framework for least-squares support vector machine classifiers, Gaussian processes, and kernel Fisher discriminant analysis. Neural Comput 14:1115–1147
19. Møller MF (1993) A scaled conjugate gradient algorithm for fast supervised learning. Neural Netw 6:525–533
20. Kulkarni AV, Drake JM, Armstrong DC, Dirks PB (1999) Measurement of ventricular size: reliability of the frontal and occipital horn ratio compared to subjective assessment. Pediatr Neurosurg 31:65–70
21. Harbaugh RE (2018) Editorial. Artificial neural networks for neurosurgical diagnosis, prognosis, and management. Neurosurg Focus 45:E3
22. Rossi NB, Khan NR, Jones TL, Lepard J, McAbee JH, Klimo P Jr (2016) Predicting shunt failure in children: should the global shunt revision rate be a quality measure? J Neurosurg Pediatr 17:249–259
23. Simon TD, Hall M, Riva-Cambrin J, Albert JE, Jeffries HE, Lafleur B, Dean JM, Kestle JR (2009) Infection rates following initial cerebrospinal fluid shunt placement across pediatric hospitals in the United States. Clinical article. J Neurosurg Pediatr 4:156–165
24. Simon TD, Kronman MP, Whitlock KB, Gove N, Browd SR, Holubkov R, Kestle JR, Kulkarni AV, Langley M, Limbrick DD Jr, Luerssen TG, Oakes J, Riva-Cambrin J, Rozzelle C, Shannon C, Tamber M, Wellons JC 3rd, Whitehead WE, Mayer-Hamblett N (2016) Variability in management of first cerebrospinal fluid shunt infection: a prospective multi-institutional observational cohort study. J Pediatr 179:185–191.e182

25. Simon TD, Riva-Cambrin J, Srivastava R, Bratton SL, Dean JM, Kestle JR (2008) Hospital care for children with hydrocephalus in the United States: utilization, charges, comorbidities, and deaths. J Neurosurg Pediatr 1:131–137
26. Spader HS, Hertzler DA, Kestle JR, Riva-Cambrin J (2015) Risk factors for infection and the effect of an institutional shunt protocol on the incidence of ventricular access device infections in preterm infants. J Neurosurg Pediatr 15:156–160
27. Whitehead WE, Riva-Cambrin J, Kulkarni AV, Wellons JC 3rd, Rozzelle CJ, Tamber MS, Limbrick DD Jr, Browd SR, Naftel RP, Shannon CN, Simon TD, Holubkov R, Illner A, Cochrane DD, Drake JM, Luerssen TG, Oakes WJ, Kestle JR (2017) Ventricular catheter entry site and not catheter tip location predicts shunt survival: a secondary analysis of 3 large pediatric hydrocephalus studies. J Neurosurg Pediatr 19:157–167
28. Zou J, Han Y, So SS (2008) Overview of artificial neural networks. Methods Mol Biol 458:15–23
29. Kousi M, Katsanis N (2016) The genetic basis of hydrocephalus. Annu Rev Neurosci 39:409–435
30. Hale AT, Wellons JC, Limbrick DD, Schiff SJ, Gamazon ER (2020) Alterations in white matter and total brain volumes underlie genetic risk of hydrocephalus. Neurosurgery 67
31. Karimy JK, Zhang J, Kurland DB, Theriault BC, Duran D, Stokum JA, Furey CG, Zhou X, Mansuri MS, Montejo J, Vera A, DiLuna ML, Delpire E, Alper SL, Gunel M, Gerzanich V, Medzhitov R, Simard JM, Kahle KT (2017) Inflammation-dependent cerebrospinal fluid hypersecretion by the choroid plexus epithelium in posthemorrhagic hydrocephalus. Nat Med 23:997–1003
32. Sullivan W, Reeves BC, Duy PQ, Nelson-Williams C, Dong W, Jin SC, Kahle KT (2020) Exome sequencing as a potential diagnostic adjunct in sporadic congenital hydrocephalus. JAMA Pediatr
33. Brookes E, Shi Y (2014) Diverse epigenetic mechanisms of human disease. Annu Rev Genet 48:237–268
34. Iakoucheva LM, Muotri AR, Sebat J (2019) Getting to the cores of autism. Cell 178:1287–1298
35. Leu C, Stevelink R, Smith AW, Goleva SB, Kanai M, Ferguson L, Campbell C, Kamatani Y, Okada Y, Sisodiya SM, Cavalleri GL, Koeleman BPC, Lerche H, Jehi L, Davis LK, Najm IM, Palotie A, Daly MJ, Busch RM, Lal D (2019) Polygenic burden in focal and generalized epilepsies. Brain 142:3473–3481
36. Peck G, Smeeth L, Whittaker J, Casas JP, Hingorani A, Sharma P (2008) The genetics of primary haemorrhagic stroke, subarachnoid haemorrhage and ruptured intracranial aneurysms in adults. PLoS One 3:e3691
37. Southerland AM, Meschia JF, Worrall BB (2013) Shared associations of nonatherosclerotic, large-vessel, cerebrovascular arteriopathies: considering intracranial aneurysms, cervical artery dissection, moyamoya disease and fibromuscular dysplasia. Curr Opin Neurol 26:13–28
38. Rudin C (2019) Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nat Mach Intell 1:206–215
39. Bydon M, Schirmer CM, Oermann EK, Kitagawa RS, Pouratian N, Davies J, Sharan A, Chambless LB (2020) Big data defined: a practical review for neurosurgeons. World Neurosurg 133:e842–e849
40. Oravec CS, Motiwala M, Reed K, Jones TL, Klimo P Jr (2019) Big data research in pediatric neurosurgery: content, statistical output, and bibliometric analysis. Pediatr Neurosurg 54:85–97
41. Pasini A (2015) Artificial neural networks for small dataset analysis. J Thorac Dis 7:953–960
42. Hale AT, Stanton AN, Zhao S, Haji F, Gannon SR, Arynchyna A, Wellons JC, Rocque BG, Naftel RP (2019) Predictors of endoscopic third ventriculostomy ostomy status in patients who experience failure of endoscopic third ventriculostomy with choroid plexus cauterization. J Neurosurg Pediatr:1–6
43. Kulkarni AV, Riva-Cambrin J, Browd SR, Drake JM, Holubkov R, Kestle JR, Limbrick DD, Rozzelle CJ, Simon TD, Tamber MS, Wellons JC 3rd, Whitehead WE (2014) Endoscopic third ventriculostomy and choroid plexus cauterization in infants with hydrocephalus: a retrospective Hydrocephalus Clinical Research Network study. J Neurosurg Pediatr 14:224–229
44. Kulkarni AV, Riva-Cambrin J, Holubkov R, Browd SR, Cochrane DD, Drake JM, Limbrick DD, Rozzelle CJ, Simon TD, Tamber MS, Wellons JC 3rd, Whitehead WE, Kestle JR (2016) Endoscopic third ventriculostomy in children: prospective, multicenter results from the Hydrocephalus Clinical Research Network. J Neurosurg Pediatr 18:423–429
45. Kulkarni AV, Riva-Cambrin J, Rozzelle CJ, Naftel RP, Alvey JS, Reeder RW, Holubkov R, Browd SR, Cochrane DD, Limbrick DD, Simon TD, Tamber M, Wellons JC, Whitehead WE, Kestle JRW (2018) Endoscopic third ventriculostomy and choroid plexus cauterization in infant hydrocephalus: a prospective study by the Hydrocephalus Clinical Research Network. J Neurosurg Pediatr 21:214–223
46. Riva-Cambrin J, Kestle JRW, Rozzelle CJ, Naftel RP, Alvey JS, Reeder RW, Holubkov R, Browd SR, Cochrane DD, Limbrick DD, Shannon CN, Simon TD, Tamber MS, Wellons JC, Whitehead WE, Kulkarni AV (2019) Predictors of success for combined endoscopic third ventriculostomy and choroid plexus cauterization in a North American setting: a Hydrocephalus Clinical Research Network study. J Neurosurg Pediatr:1–11