Prediction of Shunt Responsiveness in Suspected Patients With Normal Pressure Hydrocephalus Using the Lumbar Infusion Test: A Machine Learning Approach

Arnošt Mládek, MSc, PhD[‡§||]
Václav Gerla, MSc, PhD[¶*]
Petr Skalický, MD[‡§*]
Aleš Vlasák, MD[‡§]
Awista Zazay, BS[#]
Lenka Lhotská, MSc, PhD[¶**]
Vladimír Beneš, Sr, MD, PhD[‡]
Vladimír Beneš, Jr, MD, PhD[§]
Ondřej Bradáč, MD, MSc, PhD[‡§]

[‡]Department of Neurosurgery and Neurooncology, Military University Hospital, 1st Faculty of Medicine, Charles University in Prague, Prague, Czech Republic; [§]Department of Neurosurgery, Motol University Hospital, 2nd Faculty of Medicine, Charles University in Prague, Prague, Czech Republic; [||]Czech Technical University, Prague, Czech Republic; [¶]Department of Cognitive Systems and Neurosciences, Czech Institute of Informatics, Robotics and Cybernetics, Czech Technical University, Prague, Czech Republic; [#]Institute of Pathological Physiology, 1st Faculty of Medicine, Charles University in Prague, Prague, Czech Republic; **Department of Natural Sciences, Faculty of Biomedical Engineering, Czech Technical University, Prague, Czech Republic

*Václav Gerla and Petr Skalický contributed equally to this work.

Correspondence:
Ondřej Bradáč, MD, MSc, PhD,
Department of Neurosurgery,
Motol University Hospital,
2nd Faculty of Medicine,
Charles University in Prague,
V Úvalu 84,
Prague 5, 150 06, Czech Republic.
Email: Ondrej.Bradac2@fnmotol.cz

Received, July 3, 2021.
Accepted, October 27, 2021.
Published Online, January 28, 2022.

BACKGROUND: Machine learning (ML) approaches can significantly improve the classical R_{out}-based evaluation of the lumbar infusion test (LIT) and the clinical management of the normal pressure hydrocephalus.

OBJECTIVE: To develop a ML model that accurately identifies patients as candidates for permanent cerebral spinal fluid shunt implantation using only intracranial pressure and electrocardiogram signals recorded throughout LIT.

METHODS: This was a single-center cohort study of prospectively collected data of 96 patients who underwent LIT and 5-day external lumbar cerebral spinal fluid drainage (external lumbar drainage) as a reference diagnostic method. A set of selected 48 intracranial pressure/electrocardiogram complex signal waveform features describing nonlinear behavior, wavelet transform spectral signatures, or recurrent map patterns were calculated for each patient. After applying a leave-one-out cross-validation training–testing split of the data set, we trained and evaluated the performance of various state-of-the-art ML algorithms.

RESULTS: The highest performing ML algorithm was the eXtreme Gradient Boosting. This model showed a good calibration and discrimination on the testing data, with an area under the receiver operating characteristic curve of 0.891 (accuracy: 82.3%, sensitivity: 86.1%, and specificity: 73.9%) obtained for 8 selected features. Our ML model clearly outperforms the classical R_{out}-based manual classification commonly used in clinical practice with an accuracy of 62.5%.

CONCLUSION: This study successfully used the ML approach to predict the outcome of a 5-day external lumbar drainage and hence which patients are likely to benefit from permanent shunt implantation. Our automated ML model thus enhances the diagnostic utility of LIT in management.

KEY WORDS: Normal pressure hydrocephalus, NPH, Lumbar infusion test, LIT, Ventriculoperitoneal shunt, VP shunt, Machine learning, ICP waveform features

Neurosurgery 00:1–12, 2022 DOI:10.1227/NEU.0000000000001838 www.neurosurgery-online.com

ABBREVIATIONS: ADC, analog-to-digital converter; **AUC,** area under the curve; **AdaBoost,** adaptive boosting; **CSF,** cerebral spinal fluid; **dICP,** intracranial pressure time derivative; **ELD,** external lumbar drainage; **GaussNB,** Gaussian Naive Bayes; **GradientBoost,** gradient boosting; **ICP,** intracranial pressure; **LAM,** laminarity; **LIT,** lumbar infusion test; **LogReg,** logistic regression; **ML,** machine learning; **NPH,** normal pressure hydrocephalus; **RF,** random forest; **ROC,** receiver operating characteristic; **SVM,** support vector model; **TT,** tap test; **VP,** ventriculoperitoneal; **XGBoost,** eXtreme gradient boosting

Supplemental digital content is available for this article at www.neurosurgery-online.com.

Normal pressure hydrocephalus (NPH) is a neurodegenerative disease with a steadily rising prevalence.[1] NPH is characterized by the Hakim triad of urinary incontinence, dementia, and gait disturbance,[2] combined with ventriculomegaly, but without signs of an obstructive hydrocephalus or raised intracranial pressure (ICP). The classical "textbook" triad is, however, not always expressed and might be eclipsed by other comorbidities and signs of the natural aging process.[3]

Ventriculoperitoneal (VP) shunt insertion is the mainstay treatment for NPH [4] because it frequently halts further functional decline and

improves quality-of-life outcomes in 70% to 90% of treated patients.[5] NPH diagnosis is based on clinical status, MRI assessment, and evaluation of dynamic cerebral spinal fluid (CSF) parameters obtained using the lumbar infusion test (LIT), external lumbar drainage (ELD), or tap test (TT). Despite the progress in imaging methods, NPH diagnosis remains a challenge. According to Czepko et al,[6] up to 30% of patients meeting MRI criteria for NPH neither express Hakim triad nor display elevated CSF outflow resistance. It has been estimated that up to 80% of patients with NPH go unnoticed and receive inappropriate medical treatment.[5] The sensitivity and specificity of LIT for NPH diagnosis range between 56% to 100% and 50% to 90%, respectively.[7,8] Yet, it provides a helpful diagnostic tool to unmask the reduced CSF compliance indicative of NPH.[9-12] The temporary ELD, usually maintained for 3 to 5 days, is superior to LIT for positive predictive value (80%-100%), sensitivity (50%-100%), and specificity (60%-100%).[9] Despite being associated with significant complications in up to 3% of patients,[13] ELD features the highest prognostic accuracy of shunt responsiveness prediction and is generally recommended for NPH diagnosis.[14]

Presently, there is limited literature providing insight into ICP/CSF factors that predispose individuals toward shunt responsiveness. Knowledge of new predictive markers could be instrumental in NPH management. To explore this, we developed machine learning (ML) algorithms to predict which patients are more likely to experience clinical improvements after surgery. ML is a rapidly emerging technique with a well-documented deployment in neurosurgery[15-27]; however, the literature regarding the application of ML for diagnostics and shunt responsiveness in NPH is limited.[28-30] Similar to a simple regression model, an ML algorithm predicts an output, given a set of inputs. However, the statistical techniques behind generating the input-based predictions are more complex.

In this study, we analyzed ICP recorded throughout LIT and extracted numerous signal features of the waveform, rather than reducing the LIT outcome to one number (R_{out}) as is usually performed in practice. We then developed ML algorithms with the ability to reveal complex relations between the ICP signal features and ELD outcomes to predict which patient is more likely to respond to CSF drainage.

METHODS

IRB Statement and Guidelines

The study was conducted in accordance with the rules and regulations of our institution, as approved by the institutional ethics board. All patients signed informed consent forms before the procedures. The Transparent Reporting of Multivariable Prediction Models for Individual Prognosis or Diagnosis checklist[31] and guidelines for ML predictive models[32] were followed in our study.

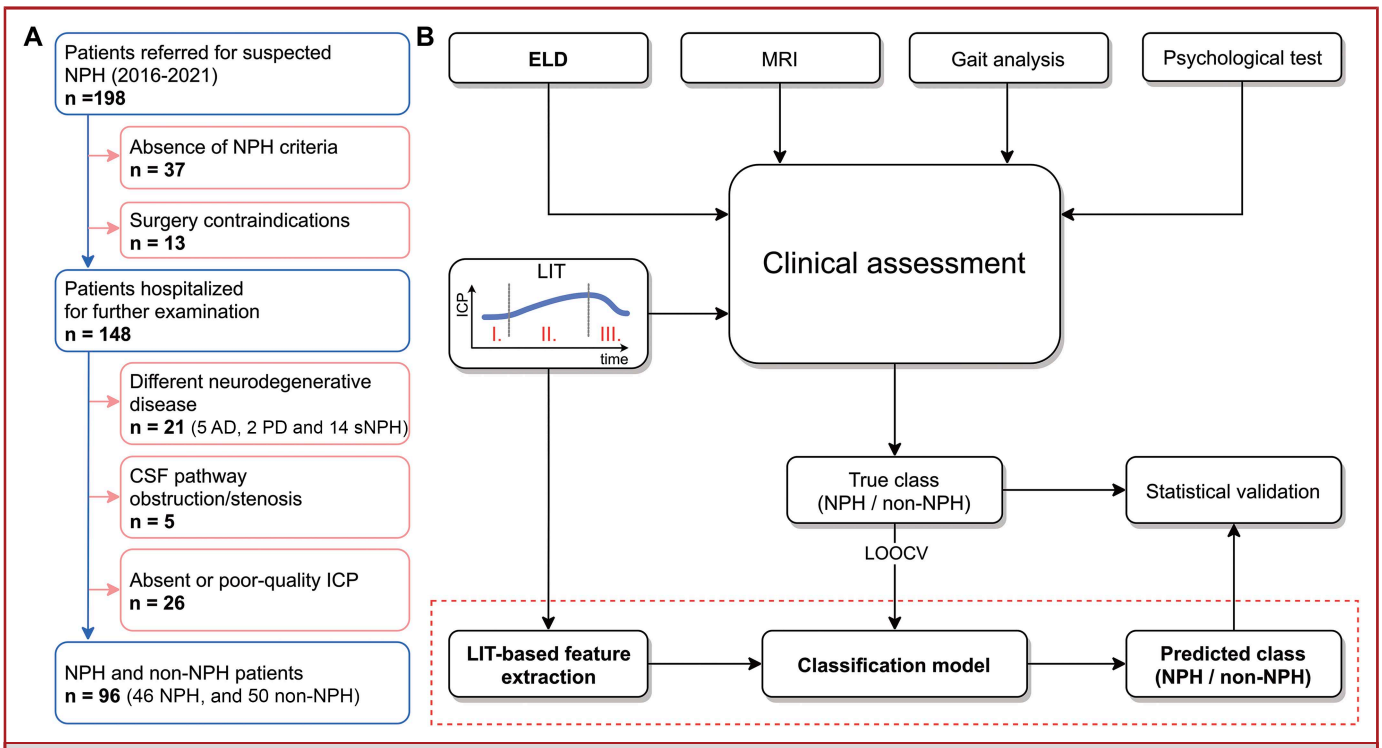

FIGURE 1. *A, The patient selection flowchart. B, Study design. ELD was used as the decisive method to classify patients as NPH (ELD+) and non-NPH (ELD−). LOOCV method of ML training process. ELD, external lumbar drainage; LOOCV, leave-one-out cross-validation; ML, machine learning, NPH, normal pressure hydrocephalus.*

TABLE 1. Description of the LIT Phases I-III

Phase	Infusion pump	Mean length	Mean ICP trend	Motive
I	Off	01:41	Steady	Equilibration of the ICP waveform, offset level
II	On (1.5 mL·min^{-1})	16:21	Increase	Main phase, testing of CSF absorptive capacity
III	Off	04:46	Decrease	Testing of the recovery rate

CSF, cerebral spinal fluid; ICP, intracranial pressure; LIT, lumbar infusion test.
The mean phase lengths are in minutes.

Design and Cohort Selection

In this single-center prospective consecutive cohort study, 198 patients with suspected NPH were referred to the Department of Neurosurgery and Neuro-oncology of the Military University Hospital in Prague between 2016 and 2021 (Figure 1A). Thirty-seven patients were excluded at the outpatient clinic because of the absence of clinical or radiological criteria for possible idiopathic NPH,[33,34] and 13 patients were excluded because of the contraindications for general anesthesia or VP shunt surgery. The remaining 148 patients were hospitalized for further diagnostic testing. All patients underwent standardized MRI and were examined by a neurologist and neuropsychologist. Twenty-one patients were excluded because of the diagnosis of another neurodegenerative disease or secondary NPH (sNPH): 5 patients with Alzheimer dementia, 2 patients with Parkinson disease, and 14 patients with sNPH. Another 5

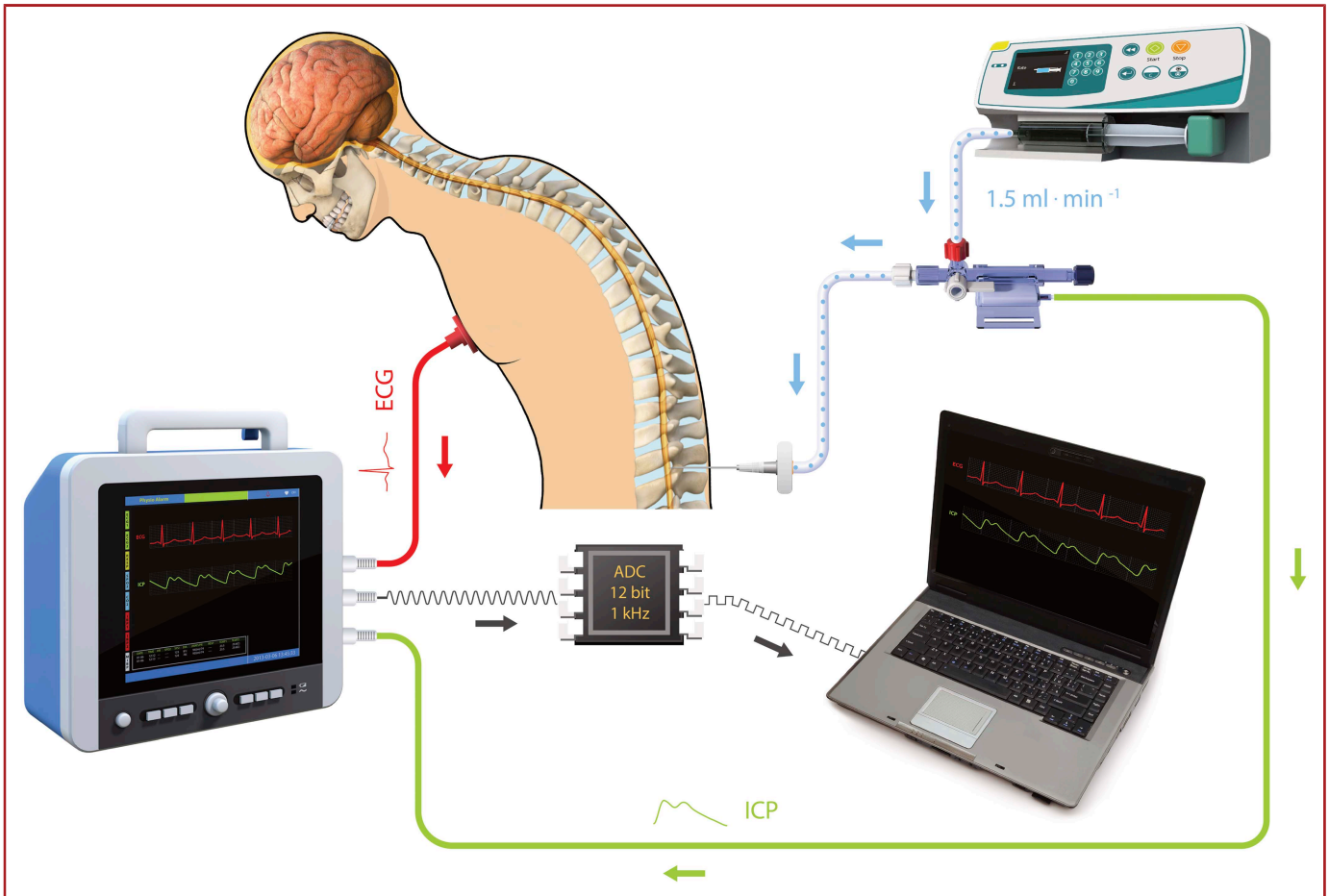

FIGURE 2. *Illustration of the LIT setup. Saline is administered by an infusion pump at a constant rate of 1.5 mL·min^{-1} into the subarachnoid space (blue arrows). The needle is connected through a fluid-filled tubing system and a 3-way stop cock to a disposable pressure transducer. The transducer is electrically connected to a bedside monitor (green curve), which displays ICP (green) and ECG (red) on-screen and simultaneously relays the analog data through ADC to a computer (gray analog and digital signal). ECG, electrocardiogram; ICP, intracranial pressure; LIT, lumbar infusion test. ADC, analog-to-digital converter.*

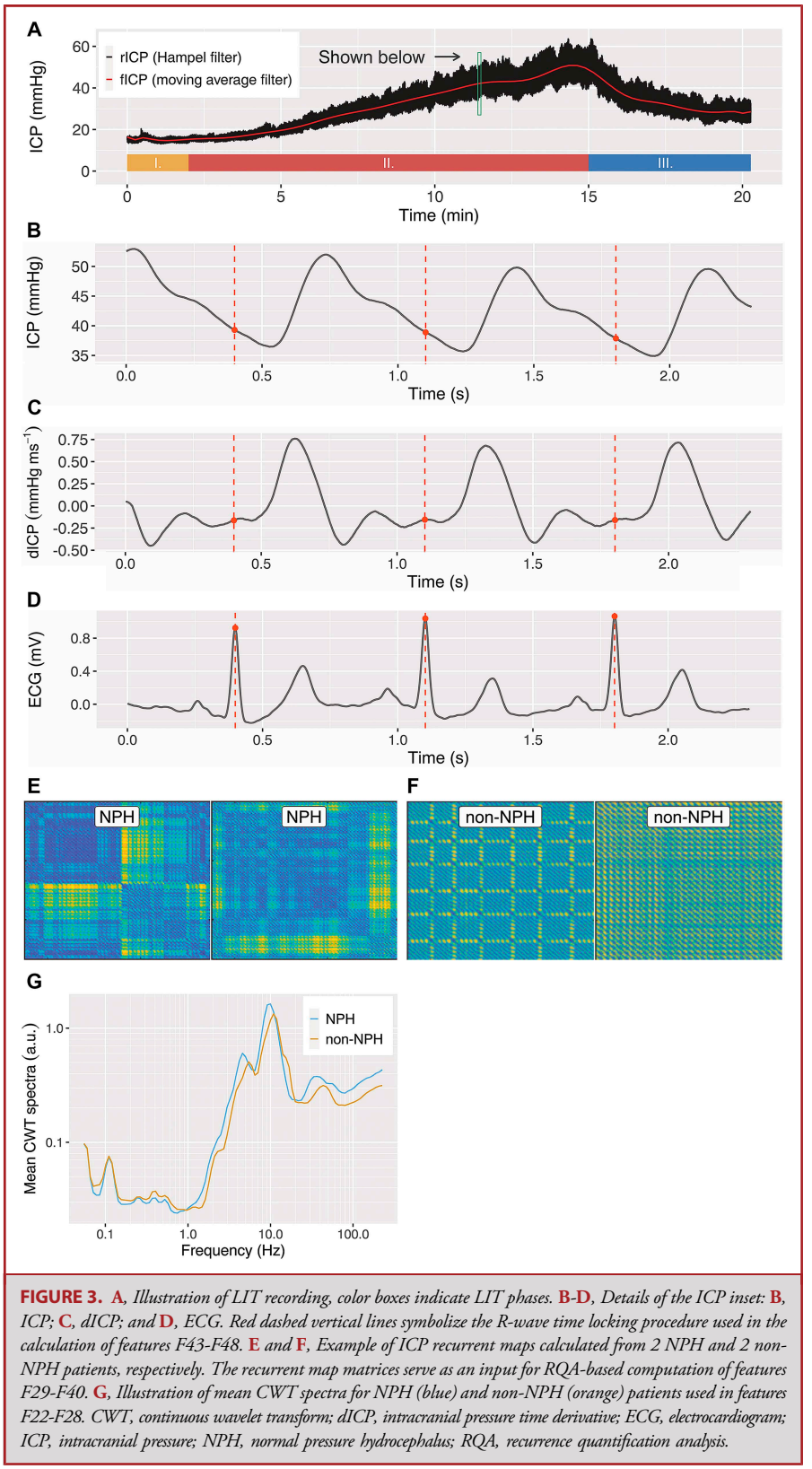

FIGURE 3. **A**, *Illustration of LIT recording, color boxes indicate LIT phases.* **B-D**, *Details of the ICP inset:* **B**, *ICP;* **C**, *dICP; and* **D**, *ECG. Red dashed vertical lines symbolize the R-wave time locking procedure used in the calculation of features F43-F48.* **E** *and* **F**, *Example of ICP recurrent maps calculated from 2 NPH and 2 non-NPH patients, respectively. The recurrent map matrices serve as an input for RQA-based computation of features F29-F40.* **G**, *Illustration of mean CWT spectra for NPH (blue) and non-NPH (orange) patients used in features F22-F28. CWT, continuous wavelet transform; dICP, intracranial pressure time derivative; ECG, electrocardiogram; ICP, intracranial pressure; NPH, normal pressure hydrocephalus; RQA, recurrence quantification analysis.*

TABLE 2. Baseline Characteristics of NPH-Suspected Patients, n = 96

Class	Variable	NPH	Non-NPH	Total	*P*-value
General	Number of patients	46 (48)	50 (52)	96 (100)	NA
	Sex (M/F)	30/16 (65/35)	35/15 (70/30)	65/31 (68/32)	.666
	Age (y)	73.5 ± 4.7	73.9 ± 8.6	73.7 ± 6.9	.803
	CCI	5.8 ± 1.8	5.8 ± 2.0	5.8 ± 1.9	.935
LIT	Positive/negative	29/17 (63/37)	19/31 (38/62)	48/48 (50/50)	.024
	Mean phase duration (I/II/III) (min)	01:46/16:13/04:35	01:36/16:28/04:56	01:41/16:21/04:46	.284/.587/.214
Hakim triads	Gait impairment (y/n/NA)	46/0/0 (100/0/0)	48/1/1 (96/2/2)	94/1/1 (98/1/1)	1.000
	Incontinence (y/n/NA)	39/7/0 (85/15/0)	31/18/1 (62/36/2)	70/25/1 (73/26/1)	.021
	Dementia (y/n/NA)	43/3/0 (93/7/0)	42/7/1 (84/14/2)	85/10/1 (89/10/1)	.319
	Patients with 3/2/1/0 signs	36/10/0/0 (78/22/0/0)	28/17/3/2 (56/34/6/4)	64/27/3/2 (67/28/3/2)	.055[a]
Postshunt improvement	Follow-up: 3 mo (y/n/NA)	33/8/5 (80/20)	NA	NA	NA
	Follow-up: 1 y (y/n/NA)	18/6/22 (75/25)	NA	NA	NA

CCI, Charlson Comorbidity Index; ELD, external lumbar drainage; NA, not applicable or available; NPH, normal pressure hydrocephalus.
[a]Freeman–Halton extension of the Fisher exact test for a 4-column contingency table was used.
Numbers in parentheses are in %. The follow-up relative numbers are calculated excluding NA. *P*-values for continuous and binary/count data were calculated by the *t*-test and Fisher exact test, respectively.

patients were diagnosed with obstruction or stenosis in the CSF pathway and underwent endoscopic third ventriculostomy. Gait was evaluated from a video recording using the Dutch Gait scale.[35,36] Twenty-six patients were excluded because of the lack of LIT or insufficient ICP recording length or quality (eg, premature termination, poor-quality recording in agitated patients, or missing electrocardiogram [ECG]). The final study group included 96 patients labeled "possible NPH." All patients with possible NPH had ventriculomegaly (Evans index >0.3) and normal CSF opening pressure.[37] The patients with possible NPH were further classified as NPH (46) and non-NPH (50) according to ELD outcomes (Figure 1B). The NPH group thus consists of patients with possible NPH with positive ELD (ELD+), and the non-NPH group consists of patients with possible NPH with negative ELD (ELD−). All patients in the NPH group underwent VP shunt implantation. Clinical outcome was assessed at 3 months and for those admitted before March 2020 again at 1 year. Non-NPH patients were referred for further neurological care. The selection for shunt implantation was based on the available guidelines.[33,37] The tap test or presence of disproportionately enlarged subarachnoid space on MRI (aka DESH; tight, high-convexity, and medial subarachnoid spaces and enlarged Sylvian fissures with ventriculomegaly) with typical gait symptoms was not considered for shunt implantation decision.[38]

Lumbar Infusion Test and External Lumbar Drainage Protocol

The LIT protocol consisted of 3 phases (Table 1 and Figure 2). A common LIT measure used to determine shunt responsiveness is resistance to CSF outflow (R_{out}; mm Hg·mL^{-1}·min)[39] calculated as the pressure gradient divided by the infusion rate (1.5 mL·min^{-1}).[40] The R_{out} ≥12 mm Hg·mL^{-1}·min was used to label the patients as LIT positive.[41] ECG was synchronously recorded for time-locking ICP segmentation only. ELD was inserted immediately after the LIT and left in place for 5 days after which ELD response was assessed. For detailed information on LIT and ELD protocols, see Supplementary Information, http://links.lww.com/NEU/B16.

ICP Feature Extractions

The extraction of ICP signal features (Figure 3) aims to uncover waveform patterns that might be specific for either of the NPH or non-NPH groups. The calculated 48 features are clustered into 7 classes depending on their nature: temporal dynamics (F01-F11), integral (F12-F13), nonlinear (F14-F21),[42-45] continuous wavelet transform (CWT; F22-F28), recurrence quantification analysis (RQA; F29-F40),[46] heart rate (HR; F41-F42), and ECG locking (F43-F48). For further description of calculated features, see Supplementary Information, http://links.lww.com/NEU/B16.

Classification Model Training and Performance Evaluation

The following ML algorithms were trained on a training set using the leave-one-out cross-validation method: random forest, logistic regression, Gaussian Naive Bayes, support vector model, adaptive boosting, extra-trees, gradient boosting, and eXtreme gradient boosting (XGBoost). The listed algorithms are implemented in the Scikit-Learn Python library.[47] Accuracy, sensitivity, specificity, receiver operating characteristic (ROC), and area under the ROC curve (area under the curve[AUC]) were used to compare the performance of all ML methods. For detailed information on ML models construction and cross-validation, see Supplementary Information, http://links.lww.com/NEU/B16.

RESULTS

Baseline Patient Data

Table 2 summarizes the baseline characteristics of the overall data. The average age of the cohort was 73.7 years. Sixty-eight percent of patients were male. Of 46 patients with NPH, 29 (63%) and 17 (37%) had positive and negative LIT, respectively. The complete Hakim triad was expressed in 36 patients with NPH (78%). In the non-NPH group, 28 patients (56%) showed the complete triad. Gait disorder as classified by the Dutch Gait Scale[35] was the most common of the triad

TABLE 3. Features Ranked According to Their Class and Feature Importance

		NPH		Non-NPH				
Feature class	**ID**	**Mean**	**SD**	**Mean**	**SD**	***P*-Value**	**FI**	**Feature description**
Temporal dynamics	**01**	**20.9**	**7.3**	**14.9**	**5.4**	**1.41e-5**	**62**	**$Q_{0.99}$(ICP) − $Q_{0.01}$(ICP); fICP**
	02	**17.7**	**6.7**	**13.0**	**4.6**	**1.32e-4**	**26**	**Mean (ICP(L_30s)) − mean (ICP(F_30s)); fICP**
	03	3.23	1.21	2.68	1.16	2.65e-2	10	SD (ICP); rICP-fICP
	04	1.22	0.63	0.83	0.29	1.83e-4	9	Mean (1 min ΔICP increments); fICP
	05	18.1	6.7	13.7	4.8	3.14e-4	8	Mean (ICP(L_60s)); fICP
	06	−2.42	1.15	−1.65	0.71	1.75e-4	6	Secant slope (ICP); LIT phase III.; fICP
	07	1.43	0.72	1.02	0.55	1.85e-3	2	SD (1 min ΔICP increments); fICP
	08	1.16	0.52	0.82	0.29	1.59e-4	2	Secant slope (ICP); fICP
	09	12.5	3.8	11.2	3.5	9.20e-2	1	Mean (ICP); LIT phase I; fICP
	10	8.57	3.29	6.92	3.23	1.50e-2	1	$Q_{0.99}$ (ICP); rICP-fICP
	11	27.8	9.1	21.1	7.3	1.30e-4	0	$Q_{0.99}$ (ICP) − $Q_{0.01}$(ICP); rICP
Integral	**12**	**1.84e+3**	**2.80e+3**	**2.54e+3**	**4.38e+3**	**3.55e-1**	**33**	**Normalized AUC; rICP-fICP**
	13	1.38e+6	3.51e+5	1.15e+6	3.26e+5	9.73e-4	4	Normalized AUC; fICP
Nonlinear	**14**	**−4.07e+7**	**3.78e+7**	**−2.96e+7**	**3.47e+7**	1.36e-**1**	**46**	**En(ICP); rICP-fICP**
	15	−7.16e+3	3.91e+3	−4.27e+3	2.48e+3	3.39e-5	12	En(ICP(L_120s)); fICP
	16	6.73	0.57	6.26	0.63	2.40e-4	4	LogEn(ICP(L_120s)); fICP
	17	1.78	0.12	1.85	0.08	1.11e-3	2	HFD(ICP(L_120s)); fICP
	18	−0.16	0.13	−0.08	0.09	2.64e-3	2	HFD(ICP(L_120s)) − HFD (ICP(F_120s)); fICP
	19	1.93	0.08	1.93	0.08	9.76e-1	1	HFD(ICP(F_120s)); fICP
	20	−1.15e+3	8.06e+2	−8.45e+2	5.12e+2	2.70e-2	1	En(ICP(F_120s)); fICP
	21	5.13	0.63	4.85	0.73	5.01e-2	0	LogEn(ICP(F_120s)); fICP
CWT-based	**22**	**0.03**	**0.01**	**0.03**	**0.01**	**4.11e-2**	**51**	**Mean, (CWT power in 0.18-0.62 Hz); rICP**
	23	0.05	0.01	0.05	0.01	9.94e-2	15	Mean, (CWT power in 0.05-0.18 Hz); rICP
	24	0.57	0.27	0.42	0.17	2.25e-3	2	Mean, (CWT power in 1.92-10.00 Hz); rICP
	25	0.03	0.01	0.03	0.01	1.60e-2	0	Mean, (CWT power in 0.62-1.92 Hz); rICP
	26	0.40	0.12	0.42	0.12	4.38e-1	0	Pos(max(CWT in 0.18-0.62 Hz)); rICP
	27	1.83	0.26	1.79	0.38	5.27e-1	0	Pos(max(CWT in 0.62-1.92 Hz)); rICP
	28	9.40	1.54	9.77	1.33	2.12e-1	0	Pos(max(CWT in 1.92-10.00 Hz)); rICP
RQA	**29**	**98.9**	**0.67**	**99.2**	**0.41**	**1.33e-1**	**31**	**DET(R(F_30s)); rICP**
	30	**99.1**	**0.7**	**99.4**	**0.3**	**4.57e-3**	**23**	**DET(R(L_30s)); rICP**
	31	**20.4**	**4.0**	**22.3**	**3.9**	**1.57e-2**	**19**	**TT(R(L_30s)); rICP**
	32	5.25	0.63	5.47	0.39	4.05e-2	12	En(R(F_30s)); rICP
	33	5.06	0.63	5.31	0.42	2.66e-2	7	En(R(mean(A_30s))); rICP
	34	99.8	0.2	99.8	0.1	1.30e-2	7	LAM(R(mean(A_30s))); rICP
	35	99.6	0.4	99.7	0.3	2.55e-1	6	LAM(R(F_30s)); rICP
	36	99.8	0.2	99.9	0.1	2.71e-2	6	LAM(R(L_30s)); rICP
	37	22.0	4.5	23.5	4.2	7.64e-2	3	TT(R(mean(A_30s))); rICP
	38	24.1	8.6	25.9	10.3	3.70e-1	2	TT(R(F_30s)); rICP
	39	4.93	0.69	5.18	0.51	4.58e-2	1	En(R(L_30s)); rICP
	40	99.1	0.6	99.3	0.3	6.94e-3	1	DET (R(mean(A_30s))); rICP
Heart rate	41	76.6	9.5	69.6	10.2	2.49e-2	2	Mean (HR); (ECG)
	42	77.0	9.1	69.8	10.7	2.26e-2	0	Median (HR); (ECG)
ECG locking–based	43	232	49	263	65	8.38e-2	14	Pos(max(dICP(F_200ms))); rICP
	44	373	70	401	74	1.99e-1	0	Pos(max(ICP(F_200ms))); rICP
	45	227	49	262	62	4.25e-2	0	Pos (max(dICP(F_500ms))); rICP
	46	230	94	249	60	4.38e-1	0	Pos (max, (dICP(L_500ms))); rICP
	47	0.30	0.13	0.18	0.07	5.35e-4	0	Pos (max(ICP(L_200ms))); rICP
	48	−0.04	0.21	−0.09	0.24	5.19e-1	0	Skew (dICP(L_200ms))–skew(dICP(F_200ms)); rICP

AUC, area under the curve; CWT, continuous wavelet transform; DET, determinism; dICP, ICP time derivative; EN, entropy; FI, XGBoost classifier feature importance; fICP, rICP signal smoothed using moving median filter with the window size of 30s; HFD, Higuchi fractal dimension; LAM, laminarity; LogEn, log energy entropy; max, maximum; min, minimum; pos, position of max/min on the time or frequency axis; Q, quantile; R, recurrent map; rICP, raw ICP filtered with the Hampel filter; rICP-fICP, filtered signal subtracted from the raw signal (offset removal); RQA, recurrence quantification analysis; SD, standard deviation; skew, skewness; TT, trapping time.

Interval of the input vector/array is tagged as X and _Y, where X = F, first; L, last, A, all segments and Y = time (eg, F_60s refers to the first 60 s of the respective variable, A_30s stands for all 30 s segments). Except for F06 and F09 calculated for LIT phases III and I, respectively, all features were computed using LIT phase II. ICP/ECG. Bold indicates 8 features with the highest FI. For detailed feature description, see Supplementary Information, http://links.lww.com/NEU/B16.

TABLE 4. Comparison of the AUCs, Accuracies, Sensitivities, and Specificities of Tested ML Models

Model	AUC	Accuracy (%)	Sensitivity (%)	Specificity (%)
Manual R_{out}-based	NA	62.5	62.0	63.0
RF	0.707	68.8	72.0	63.0
LogReg	0.711	70.8	80.0	60.9
GaussNB	0.688	71.6	84.0	52.2
SVM	0.728	71.9	86.0	56.5
AdaBoost	0.707	75.0	84.0	65.2
ExtraTrees	0.817	76.0	82.0	69.6
GradientBoost	0.895	79.2	86.0	71.7
XGBoost	0.887/0.891 (8)	80.2/82.3 (8)	86.0/86.1 (8)	73.9/78.3 (7)

AUC, area under the curve; AdaBoost, adaptive boosting; ExtraTrees, extra-trees; GaussNB, Gaussian Naive Bayes; GradientBoost, gradient boosting; LogReg, logistic regression; ML, machine learning; NA, not available; RF, random forest; SVM, support vector model; XGBoost, eXtreme gradient boosting.
In XGBoost X/Y (Z): X refers to the performance obtained for all 48 features, Y represents the highest performance obtained for optimal number of features, and Z is the optimal feature number.

signs seen in nearly all patients. The incidence of urinary incontinence was the lowest. There were no significant differences in sex, age, and number of patients between the NPH and non-NPH groups (Table 2).

The 3-month follow-up of the shunted patients revealed improvement in 80% and no significant amelioration in 20%; 5 patients were not examined.

Feature Selection

Table 3 summarizes the results of the computed ICP/ECG features F01-F48 that were ultimately used to develop ML models. For detailed information on the features, see Supplementary Information, http://links.lww.com/NEU/B16.

Model Performance

Table 4 compares accuracies (Figure 4), AUCs, sensitivities, and specificities for all ML algorithms developed. From these algorithms, the XGBoost classifier showed the best discrimination potential, with 80.2% accuracy, 0.887 AUC, 86.0% sensitivity, and 73.9% specificity when all 48 features are considered. The manual R_{out}-based classification displays significantly lower concordance with ELD outcomes with accuracy, sensitivity, and specificity of 62.5%, 62.0%, and 63.0%, respectively. Because of its superior discrimination, balanced accuracy, and the ability to reveal complex feature dependencies, the XGBoost classifier was selected for calibration and further testing. Figure 5 shows the detailed performance of the XGBoost classifier. For AUC (0.891), accuracy (82.3%), and sensitivity (86.1%), the highest predictive

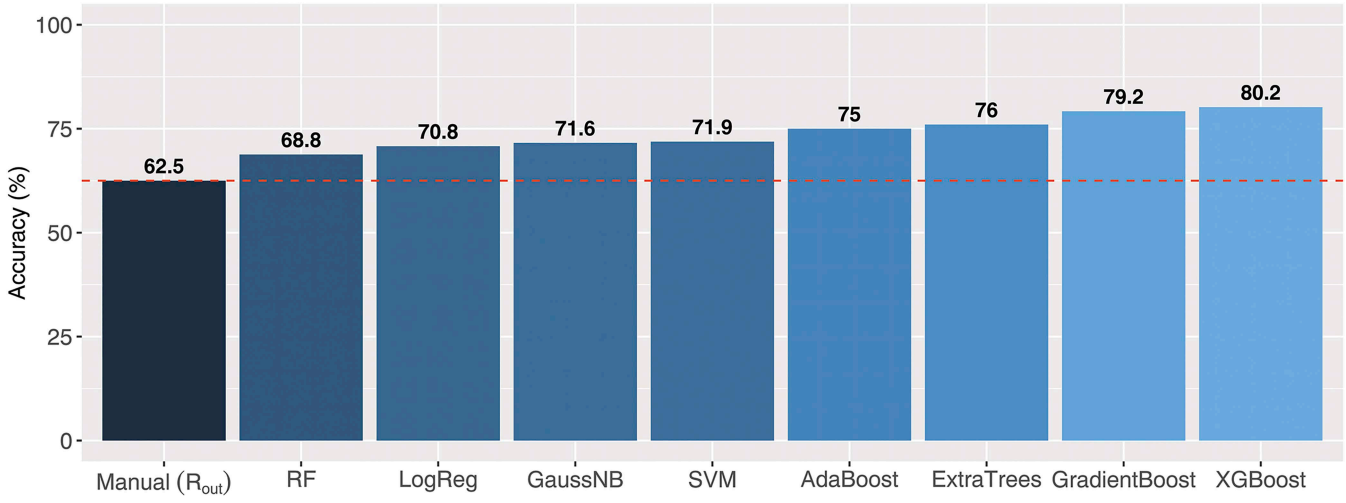

FIGURE 4. *Tested classification models. The best performance for accuracy (%) is obtained by the XGBoost technique. The red dashed line symbolizes the accuracy of the manual R_{out}-based classification. In training of ML models, the same set of 48 features were used. ML, machine learning; XGBoost, eXtreme gradient boosting.*

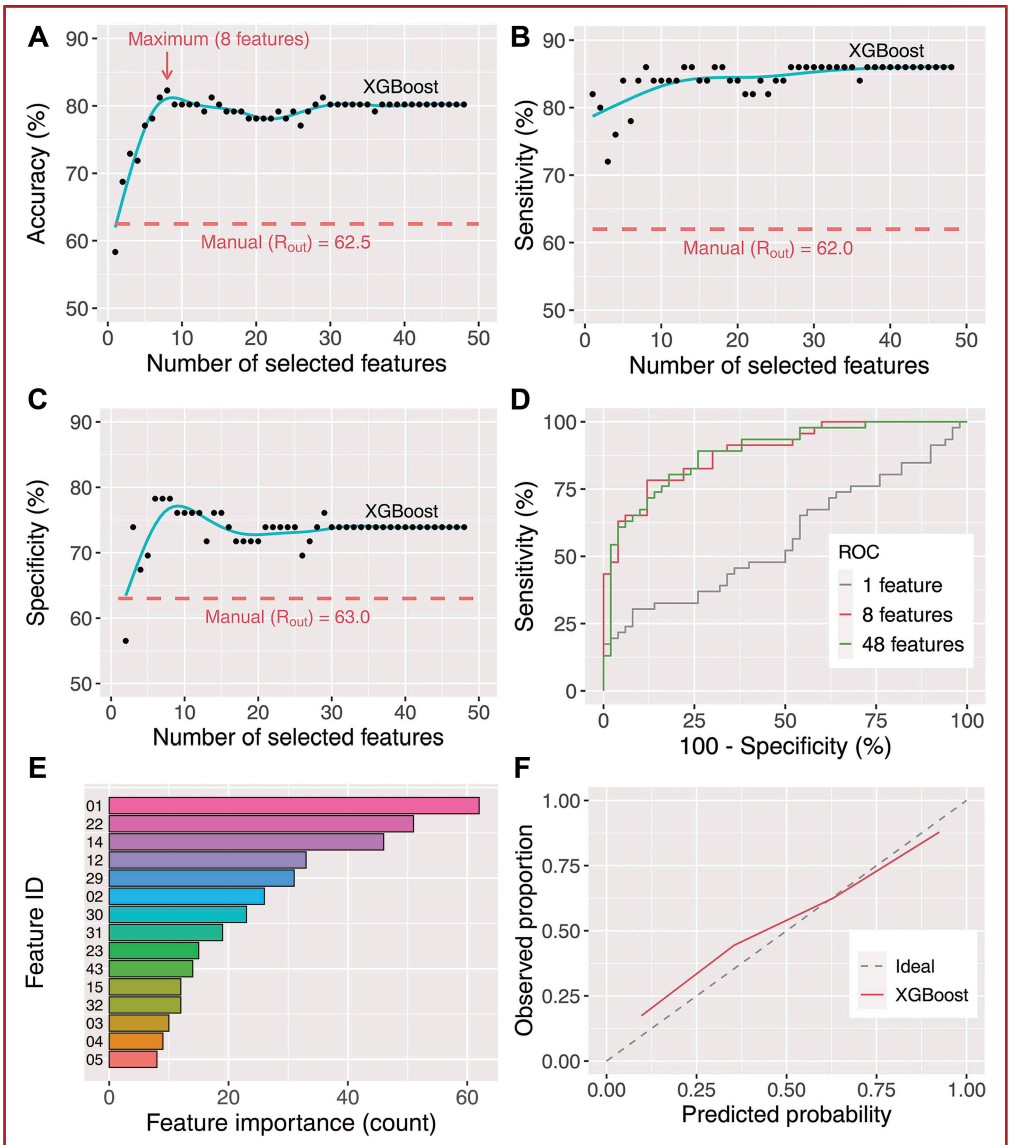

FIGURE 5. *XGBoost model performance.* **A-C**, *Red dashed lines symbolize the performance of R_{out}-based manual classification.* **D**, *ROC curves for 1, 8, and 48 features.* **E**, *Ranking the calculated features according to their importance in the model.* **F**, *Calibration curve of the XGBoost model. The gray dashed line denotes ideal calibration. XGBoost, eXtreme gradient boosting, ROC, receiver operating characteristic.*

potential was obtained for 8 features (Figure 5A and 5B). The highest specificity (78.3%) was achieved with 7 features (Figure 5C). Figure 5D shows the XGBoost model ROC curve when 1, 8, or all 48 features are considered. The feature importance (Figure 5E) indicates the most seminal predictors for NPH/non-NPH discrimination, with relative importance ranking based on usage frequency in the model. The distribution of the 8 most important predictive factors (F01, 22, 14, 12, 29, 2, 30, and 31; Table 3) in NPH and non-NPH cohorts is shown in Figure 6. The calibration curve (Figure 5F) demonstrates good concordance between the estimated and observed probabilities.

DISCUSSION

ICP waveform exhibits complex time-domain and frequency-domain motifs, knowledge of which may support clinical decision making. Although fundamental, the feature selection is of limited value when the features are used separately. Unsurprisingly, the

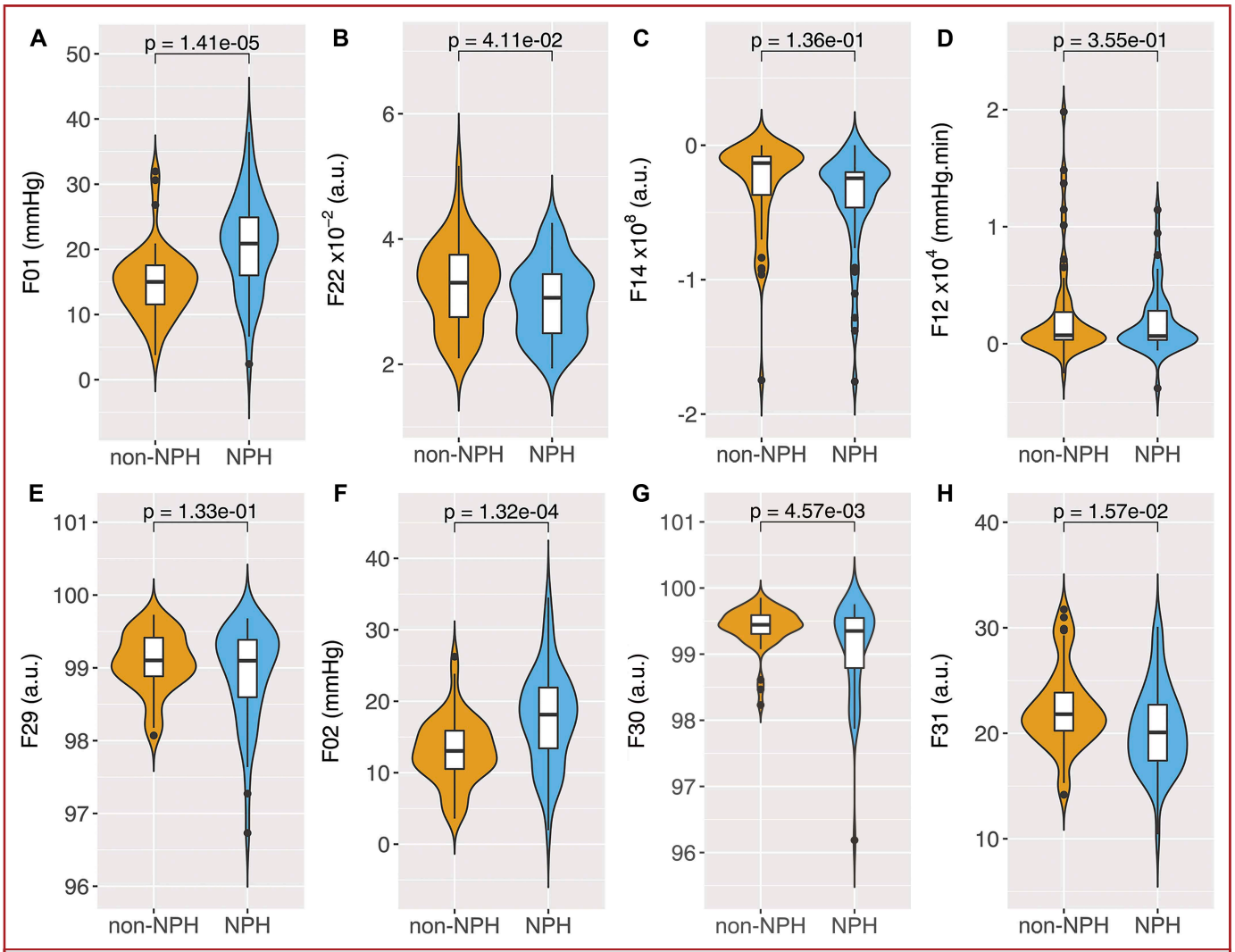

FIGURE 6. A-H, *Distribution of the 8 most used XGBoost features in NPH (=ELD+, blue) and non-NPH (=ELD−, orange) cohorts ranked according to the FI metric. ELD, external lumbar drainage; NPH, normal pressure hydrocephalus; XGBoost, eXtreme gradient boosting.*

principal feature with the highest FI is F01 (Figure 5), the ICP elevation in the phase II, which is basically the unscaled R_{out} value. F01 is significantly higher in patients with NPH (P = 1.4e-5, FI = 62, Table 3 and Figure 6), in line with the original principle of LIT. The accuracy of a potential 1-feature–based classification model (<60%, Figure 5) is slightly lower than the manual R_{out} method. This discrepancy is because the examining physician takes into account the ICP artifacts precipitated by the patient's movement, cough, etc. in the LIT evaluation while the ML model lacks this information. The remaining features F02-F48 exhibit 2 phenomenas. First, they are difficult to interpret because they lack a clear clinical correlate, and their physiological explanation is rather speculative or under further investigation. Second, FI values do not necessarily correlate with the P-values. Features that would have been ignored in standard statistical testing as insignificant for NPH/non-NPH discrimination may turn out to be seminal for the ML model and vice versa (Figure 6). While the second most used feature F22, a mean CWT spectral power within the range of 0.18–0.62 Hz (P = 4.1e-2, FI = 51), is borderline from this perspective, the third and fourth most important features F14, the entropy of the rICP-fICP difference (FI = 46, P = 1.4e-1), and F12, the integral of the normalized ICP curve (FI = 33, P = 3.6e-1), would be rejected by the significance testing. Despite being often unexplainable, the additional features F02-F48 boost our XGBoost model to enhance the prediction accuracy by ~25% compared with the F01-only predictor and by ~20% compared with the actual R_{out}-based manual evaluation. This finding illustrates the greatest asset of the ML algorithms—the ability to explore complex multidimensional feature space and reveal clinically exploitable information hidden within, unreachable through common statistical techniques.

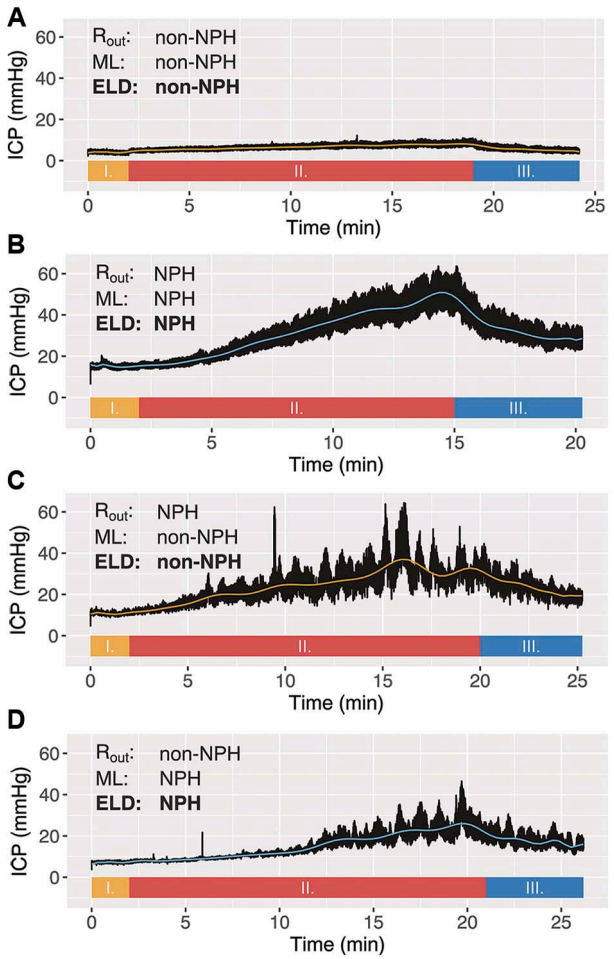

FIGURE 7. *LIT ICP of 4 selected patients A-D, P1-P4, non-NPH (orange; A, P1 and C, P3), and NPH (blue; B, P2 and D, P4). While A, P1 and B, P2 are representatives of patients where the manual R_{out}-based classification correctly discriminates between NPH and non-NPH, C, P3 and D, P4 illustrate those gray zone patients in which R_{out}-based classification fails. Specifically, R_{out} of C, P3 and D, P4 is 13.8 and 7.3 mm Hg·mL^{-1}·min, respectively, assigning them to the opposite class. A-D, All P1-P4 patients were classified in accordance with ELD using our XGBoost ML model. ELD, external lumbar drainage; ML, machine learning; NPH, normal pressure hydrocephalus; XGBoost, eXtreme gradient boosting.*

Clearly, ML-based evaluation of LIT cannot, at least for now, replace ELD completely. Still, an in-depth LIT ICP analysis may reveal a subset of patients who could be indicated for permanent CSF drainage. The benefit of our ML model is illustrated in Figure 7. Although patients P1-P2 show an easy-to-recognize ICP elevation typical for non-NPH and NPH diagnosis, patients P3-P4 represent the gray zone patients. Because R_{out} of P3 and P4 was 13.8 and 7.3 mm Hg·mL^{-1}·min, respectively, P3 was labeled LIT+ and P4 LIT−. Patient P4, however, turned out to be ELD+ and improved significantly on VP shunt insertion. Patient P3 did not show any clinical amelioration throughout ELD. Unlike the R_{out}-based assessment that failed for P3-P4, our ML model correctly predicted the ELD outcome in all patients (P1-P4).

Limitations

There are several limitations to the work presented in this study, with the selection bias being the most significant one. First, before referral for functional testing, patients are selected based on clinical and radiological criteria.[37] This selection process prematurely screens out potential ELD+ patients who would benefit from CSF diversion. Second, for a definite NPH diagnosis, shunt response must be observed.[37] Yet, ELD remains the most pragmatic test available for selecting patients for shunt insertion.[48,49] Because the specificity of ELD is generally below 100%,[9] a fraction of ELD+ patients do not show postsurgery improvement. In this study, clinical improvement at 3 and 12 months was 80% and 75% (Table 2), respectively, which is in accordance with reported data.[49] Variability of the reported values could be explained by nonidentical improvement cutoffs and different shunt indication criteria being used throughout different centers.[9] Similarly, subtotal sensitivity of ELD causes that an unknown number of ELD− patients who might have benefited from shunting are classified as non-NPH and as such do not undergo surgery. Because of the aforementioned, choosing ELD as the ultimate NPH test unavoidably leads to an inherent selection bias. Being aware of the bias, we have opted for the most pragmatic method of functional NPH diagnostics based on ELD response, and thus, we have performed shunt implantation only in patients showing significant post-ELD improvement. Functional testing before shunt implantation is recognized in many studies[9,14] and guidelines[37] and has a long tradition at our institution. If we shift our threshold for improvement after ELD to lower limits, it would inevitably lead to unnecessary surgeries in a number of patients. From this point of view, it is rather an ethical question of whether to implant a shunt system in borderline patients. Finally, patients present in various stages of the disease with frequent comorbidities that potentially influence the outcome of the functional testing.[50,51]

The abovementioned factors thus compromise the process of NPH/non-NPH classification, with no ultimate solution in sight. Our ML models were optimized for highly accurate prediction rather than explanation, and model parameters thus cannot be simply deployed for the purpose of explaining the effect of individual features on the shunt responsiveness. Finally, although the model performed well throughout the leave-one-out cross-validation, a technique proven to be useful for assessing model effectiveness,[52-54] further external validation is under way in multiple neurosurgical centers.

CONCLUSION

ML algorithms are a promising tool for prediction of postoperative outcomes, and these algorithms can be integrated into

clinical management. We retrospectively applied automated ML to predict ELD and permanent shunt implantation response in patients who underwent LIT. A XGBoost model had the best performance and showed outstanding discrimination and calibration on our data set. Our developed classifier may be a considerable asset in LIT evaluation and in identifying patients with NPH who benefit from surgical intervention. Multicentric validation of our ML model has been initiated.

Funding

Research reported in this publication was supported by the Charles University Grant Agency (GAUK, no. 1068120) and the Student Grant Competition of the Czech Technical University (SGS21/162/OHK4/3T/13).

Disclosures

The authors have no personal, financial, or institutional interest in any of the drugs, materials, or devices described in this article.

REFERENCES

1. Andersson J, Rosell M, Kockum K, Lilja-Lund O, Söderström L, Laurell K. Prevalence of idiopathic normal pressure hydrocephalus: a prospective, population-based study. *PLoS One*. 2019;14(5):e0217705.
2. Hakim S, Adams RD. The special clinical problem of symptomatic hydrocephalus with normal cerebrospinal fluid pressure. Observations on cerebrospinal fluid hydrodynamics. *J Neurol Sci.* 1965;2(4):307-327.
3. Hashimoto M, Ishikawa M, Mori E, Kuwana N; Study of INPH on neurological improvement (SINPHONI). Diagnosis of idiopathic normal pressure hydrocephalus is supported by MRI-based scheme: a prospective cohort study. *Cerebrospinal Fluid Res.* 2010;7:18.
4. Reddy GK, Bollam P, Caldito G. Long-term outcomes of ventriculoperitoneal shunt surgery in patients with hydrocephalus. *World Neurosurg.* 2014;81(2):404-410.
5. Kiefer M, Unterberg A. The differential diagnosis and treatment of normal-pressure hydrocephalus. *Dtsch Arztebl Int.* 2012;109(1-2):15-26.
6. Czepko R, Cieslicki K. Repeated assessment of suspected normal pressure hydrocephalus in non-shunted cases. A prospective study based on the constant rate lumbar infusion test. *Acta Neurochir (Wien)*. 2016;158(5):855-863.
7. Hebb AO, Cusimano MD. Idiopathic normal pressure hydrocephalus: a systematic review of diagnosis and outcome. *Neurosurgery*. 2001;49(5):1166-1184.
8. Skalický P, Mládek A, Vlasák A, De Lacy P, Beneš V, Bradáč O. Normal pressure hydrocephalus-an overview of pathophysiological mechanisms and diagnostic procedures. *Neurosurg Rev.* 2020;43(6):1451-1464.
9. Marmarou A, Bergsneider M, Klinge P, Relkin N, Black PM. The value of supplemental prognostic tests for the preoperative assessment of idiopathic normal-pressure hydrocephalus. *Neurosurgery*. 2005;57(3 suppl):S17-S28.
10. Ryding E, Kahlon B, Reinstrup P. Improved lumbar infusion test analysis for normal pressure hydrocephalus diagnosis. *Brain Behav.* 2018;8(11):e01125.
11. Kahlon B, Sundbärg G, Rehncrona S. Comparison between the lumbar infusion and CSF tap tests to predict outcome after shunt surgery in suspected normal pressure hydrocephalus. *J Neurol Neurosurg Psychiatry*. 2002;73(6):721-726.
12. Kahlon B, Sundbärg G, Rehncrona S. Lumbar infusion test in normal pressure hydrocephalus. *Acta Neurol Scand.* 2005;111(6):379-384.
13. Governale LS, Fein N, Logsdon J, Black PM. Techniques and complications of external lumbar drainage for normal pressure hydrocephalus. *Neurosurgery*. 2008; 63:379-384.
14. Walchenbach R, Geiger E, Thomeer RT, Vanneste JA. The value of temporary external lumbar CSF drainage in predicting the outcome of shunting on normal pressure hydrocephalus. *J Neurol Neurosurg Psychiatry*. 2002;72(5):503-506.
15. Yasaka K, Abe O. Deep learning and artificial intelligence in radiology: current applications and future directions. *PLoS Med.* 2018;15(11):e1002707.
16. Mei J, Desrosiers C, Frasnelli J. Machine learning for the diagnosis of Parkinson's disease: a review of literature. *Front Aging Neurosci.* 2021;13:633752.
17. Senders JT, Zaki MM, Karhade AV, et al. An introduction and overview of machine learning in neurosurgical care. *Acta Neurochir (Wien)*. 2018;160(1):29-38.
18. Segato A, Marzullo A, Calimeri F, De Momi E. Artificial intelligence for brain diseases: a systematic review. *APL Bioeng.* 2020;4(4):041503.
19. Deo RC. Machine learning in medicine. *Circulation*. 2015;132(20):1920-1930.
20. Buchlak QD, Esmaili N, Leveque JC, et al. Machine learning applications to clinical decision support in neurosurgery: an artificial intelligence augmented systematic review. *Neurosurg Rev.* 2020;43(5):1235-1253.
21. Staartjes VE, Stumpo V, Kernbach JM, et al. Machine learning in neurosurgery: a global survey. *Acta Neurochir (Wien)*. 2020;162(12):3081-3091.
22. Celtikci E. A systematic review on machine learning in neurosurgery: the future of decision-making in patient care. *Turk Neurosurg.* 2018;28(2):167-173.
23. Booth TC, Williams M, Luis A, Cardoso J, Ashkan K, Shuaib H. Machine learning and glioma imaging biomarkers. *Clin Radiol.* 2020;75(1):20-32.
24. Savarraj JPJ, Hergenroeder GW, Zhu L, et al. Machine learning to predict delayed cerebral ischemia and outcomes in subarachnoid hemorrhage. *Neurology*. 2021; 96(4):e553-e562.
25. Rau A, Kim S, Yang S, et al. SVM-based normal pressure hydrocephalus detection. *Clin Neuroradiol.* 2021;31(4):1029-1035.
26. Shao M, Han S, Carass A, et al. Brain ventricle parcellation using a deep neural network: application to patients with ventriculomegaly. *Neuroimage Clin.* 2019;23:101871.
27. Maass F, Michalke B, Willkommen D, et al. Elemental fingerprint: reassessment of a cerebrospinal fluid biomarker for Parkinson's disease. *Neurobiol Dis.* 2020;134: 104677.
28. Santamarta D, González-Martínez E, Fernández J, Mostaza A. The prediction of shunt response in idiopathic normal-pressure hydrocephalus based on intracranial pressure monitoring and lumbar infusion. *Acta Neurochir Suppl.* 2016;122: 267-274.
29. Muscas G, Matteuzzi T, Becattini E, et al. Development of machine learning models to prognosticate chronic shunt-dependent hydrocephalus after aneurysmal subarachnoid hemorrhage. *Acta Neurochir (Wien)*. 2020;162(12):3093-3105.
30. Klimont M, Flieger M, Rzeszutek J, Stachera J, Zakrzewska A, Jończyk-Potoczna K. Automated ventricular system segmentation in paediatric patients treated for hydrocephalus using deep learning methods. *Biomed Res Int.* 2019;2019: 3059170.
31. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual Prognosis or diagnosis (TRIPOD): the TRIPOD statement. *Ann Intern Med.* 2015;162(1):55-63.
32. Luo W, Phung D, Tran T, et al. Guidelines for developing and reporting machine learning predictive models in biomedical research: a multidisciplinary view. *J Med Internet Res.* 2016;18(12):e323.
33. Relkin N, Marmarou A, Klinge P, et al.. Diagnosing idiopathic normal-pressure hydrocephalus. *Neurosurgery*. 2005;57(3 suppl):S4-S16.
34. Mori E, Ishikawa M, Kato T, et al. Guidelines for management of idiopathic normal pressure hydrocephalus: second edition. *Neurol Med Chir (Tokyo)*. 2012; 52(11):775-809.
35. Ravdin LD, Katzen HL, Jackson AE, Tsakanikas D, Assuras S, Relkin NR. Features of gait most responsive to tap test in normal pressure hydrocephalus. *Clin Neurol Neurosurg.* 2008;110(5):455-461.
36. Boon AJ, Tans JT, Delwel EJ, et al. Dutch Normal-Pressure Hydrocephalus Study: randomized comparison of low- and medium-pressure shunts. *J Neurosurg.* 1998, 88(3):490-495.
37. Nakajima M, Yamada S, Miyajima M, et al. Guidelines for management of idiopathic normal pressure hydrocephalus (third edition): endorsed by the Japanese society of normal pressure hydrocephalus. *Neurol Med Chir (Tokyo)*. 2021;61(2): 63-97.
38. Craven CL, Toma AK, Mostafa T, Patel N, Watkins LD. The predictive value of DESH for shunt responsiveness in idiopathic normal pressure hydrocephalus. *J Clin Neurosci.* 2016;34:294-298.
39. Meier U, Bartels P. The importance of the intrathecal infusion test in the diagnostic of normal-pressure hydrocephalus. *Eur Neurol*. 2001;46(4):178-186.
40. Børgesen SE, Gjerris F. Relationships between intracranial pressure, ventricular size, and resistance to CSF outflow. *J Neurosurg.* 1987;67(4):535-539.
41. Kim DJ, Kim H, Kim YT, et al. Thresholds of resistance to CSF outflow in predicting shunt responsiveness. *Neurol Res.* 2015;37(4):332-340.
42. Higuchi T. Approach to an irregular time series on the basis of the fractal theory. *Phys Nonlinear Phenom.* 1988;31(2):277-283.

43. Shannon CE. A mathematical theory of communication. *Bell Syst Tech J*. 1948; 27(3):379-423.
44. Dai H, Jia X, Pahren L, Lee J, Foreman B. Intracranial pressure monitoring signals after traumatic brain injury: a narrative overview and conceptual data science framework. *Front Neurol*. 2020;11:959.
45. Esteller R, Vachtsevanos G, Echauz J, Litt B. A comparison of waveform fractal dimension algorithms. *IEEE*. 2001;48(2):177-183.
46. Yang H. Multiscale recurrence quantification analysis of spatial cardiac vectorcardiogram signals. *IEEE Trans Biomed Eng*. 2011;58(2):339-347.
47. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in Python. *J Mach Learn Res*. 2011;12:2825-2830.
48. Chotai S, Medel R, Herial NA, Medhkour A. External lumbar drain: a pragmatic test for prediction of shunt outcomes in idiopathic normal pressure hydrocephalus. *Surg Neurol Int*. 2014;5(1):12.
49. Giordan E, Palandri G, Lanzino G, Murad MH, Elder BD. Outcomes and complications of different surgical treatments for idiopathic normal pressure hydrocephalus: a systematic review and meta-analysis. *J Neurosurg*. 2019;131:1024-1036.
50. Malm J, Graff-Radford NR, Ishikawa M, et al. Influence of comorbidities in idiopathic normal pressure hydrocephalus — research and clinical care. A report of the ISHCSF task force on comorbidities in INPH. *Fluids Barriers CNS*. 2013; 10(1):22.
51. Krahulik D, Vaverka M, Hrabalek L, et al. Ventriculoperitoneal shunt in treating of idiopathic normal pressure hydrocephalus-single-center study. *Acta Neurochir (Wien)*. 2020;162(1):1-7.
52. Molinaro AM, Simon R, Pfeiffer RM. Prediction error estimation: a comparison of resampling methods. *Bioinformatics*. 2005;21(15):3301-3307.
53. Wong T-T. Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation. *Pattern Recognit*. 2015;48(9):2839-2846.
54. Arlot S, Celisse A. A survey of cross-validation procedures for model selection. *Stat Surv*. 2010;4:40-79.

COMMENT

This interesting study used a ML protocol to attempt to improve the selection of patients for VP shunting for normal pressure hydrocephalus. The results suggest that the XGBoost algorithm can predict the outcome of a prolonged external drainage trial and thus those patients who should respond favorably to creation of a permanent shunt. The automated ML protocol also improves the value of the LIT for this purpose. The notion of applying mathematical rigor to the diagnosis and treatment of neurological disease is extremely attractive. These results are an encouraging step that suggests that this goal may be achievable. Still, the authors did not end up shunting all patients who had at least some clinical indicators of NPH based on ML criteria. There are likely patients in this group who would have benefited from shunting, but the decision to do so is often a subjective one. Further still, time will tell whether the durability of these results will be superior to other selection methods.

Joel D. MacDonald
Salt Lake City, Utah, USA

Supplemental digital content *is available for this article at www.neurosurgery-online.com.*

Supplementary Information. LIT protocol details, ELD protocol details, ICP/ECG signal features: calculation and details, ML model parameters, and cross-validation scenario details. ELD, external lumbar drainage; ICP, intracranial pressure; LIT, lumbar infusion test; ML, machine learning.