

UCLA

UCLA Previously Published Works

Title

Machine learning predicts cerebral vasospasm in patients with subarachnoid haemorrhage.

Permalink

<https://escholarship.org/uc/item/12m905sn>

Authors

Zarrin, David

Suri, Abhinav

McCarthy, Karen

et al.

Publication Date

2024-07-01

DOI

10.1016/j.ebiom.2024.105206

Peer reviewed

Machine learning predicts cerebral vasospasm in patients with subarachnoid haemorrhage



David A. Zarrin,^a Abhinav Suri,^a Karen McCarthy,^b Bilwaj Gaonkar,^c Bayard R. Wilson,^c Geoffrey P. Colby,^c Robert E. Freundlich,^b and Eilon Gabel^{d,*}

^aDavid Geffen School of Medicine at University of California, Los Angeles, USA

^bDepartment of Anesthesiology, Vanderbilt University Medical Center, USA

^cDepartment of Neurological Surgery at University of California, Los Angeles Health, USA

^dDepartment of Anesthesia and Perioperative Medicine at University of California, Los Angeles Health, USA



Summary

Background Cerebral vasospasm (CV) is a feared complication which occurs after 20–40% of subarachnoid haemorrhage (SAH). It is standard practice to admit patients with SAH to intensive care for an extended period of resource-intensive monitoring. We used machine learning to predict CV requiring verapamil (CVRV) in the largest and only multi-center study to date.

Methods Patients with SAH admitted to UCLA from 2013 to 2022 and a validation cohort from VUMC from 2018 to 2023 were included. For each patient, 172 unique intensive care unit (ICU) variables were extracted through the primary endpoint, namely first verapamil administration or no verapamil. At each institution, a light gradient boosting machine (LightGBM) was trained using five-fold cross validation to predict the primary endpoint at various hospitalization timepoints.

Findings A total of 1750 patients were included from UCLA, 125 receiving verapamil. LightGBM achieved an area under the ROC (AUC) of 0.88 > 1 week in advance and ruled out 8% of non-verapamil patients with zero false negatives. Our models predicted “no CVRV” vs “CVRV within three days” vs “CVRV after three days” with AUCs = 0.88, 0.83, and 0.88, respectively. From VUMC, 1654 patients were included, 75 receiving verapamil. VUMC predictions averaged within 0.01 AUC points of UCLA predictions.

Interpretation We present an accurate and early predictor of CVRV using machine learning with multi-center validation. This represents a significant step towards optimized clinical management and resource allocation in patients with SAH.

Funding Robert E. Freundlich is supported by National Center for Advancing Translational Sciences federal grant UL1TR002243 and National Heart, Lung, and Blood Institute federal grant K23HL148640; these funders did not play any role in this study. The National Institutes of Health supports Vanderbilt University Medical Center which indirectly supported these research efforts. Neither this study nor any other authors personally received financial support for the research presented in this manuscript. No support from pharmaceutical companies was received.

Copyright Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Keywords: Cerebral vasospasm; Verapamil; Machine learning; Prediction

Introduction

Cerebral vasospasm (CV) is a common angiographic finding following subarachnoid haemorrhage (SAH) and is widely reported as a primary contributor to delayed cerebral ischemia (DCI) and concomitant morbidity and mortality in this population.¹ CV manifests with variable severity, being angiographically appreciable in up to 70% of all patients with SAH and

clinically symptomatic in 20–40% of patients with SAH.^{2,3} A smaller subset of up to 20% of all patients with CV suffer from severe CV which results in either death or severe neurologic deficit.⁴ Intraarterial verapamil, a calcium channel blocker, is the only targeted therapy available for vasospasm reversal and is therefore commonly given. Owing to the prevalence of severe CV after SAH, it is standard clinical practice to admit

eBioMedicine

2024;105: 105206

Published Online xxx

<https://doi.org/10.1016/j.ebiom.2024.105206>

1016/j.ebiom.2024.105206

105206

*Corresponding author. University of California, 10833 Le Conte Avenue, Los Angeles, CA, 90095, USA.

E-mail address: egabel@mednet.ucla.edu (E. Gabel).

Research in context

Evidence before this study

We searched PubMed for studies on cerebral vasospasm (CV) prediction published before February 1st, 2024. Search criteria were (predict*[Title]) AND (vasospasm [Title]). The references of relevant articles were also reviewed. Many attempts to predict and associate CV have been made, only a handful of which utilize machine learning (ML). The only ML attempts to achieve predictive power sufficient for clinical use were validated over small cohorts, not externally validated, leveraged a small number of clinical predictors, and often relied on labor-intensive, subjective, and inconsistently available human-generated clinical scoring systems based on radiography. The feasibility of predicting CV with ML remains to be rigorously investigated.

Added value of this study

We demonstrate that ML predicts CV requiring verapamil (CVRV) in patients with subarachnoid haemorrhage (SAH) when trained on a vast collection (172 unique) intensive care unit (ICU) datapoints. Our model achieved highly accurate (AUC = 0.88) predictions across a large cohort at our institution. Notably, our model ruled-out verapamil in an average of 8% of patients not requiring verapamil up to 10 days advance, without any false negatives. We then externally validated our model at a separate institution located across the country in a cohort of patients with SAH of nearly equal

size and found that predictive models performed nearly identically. In our validation, we restricted model inputs to only those which are routinely collected in highly standardized manner across institutions to further improve the generalizability of these findings. Throughout this study, we extensively explored ML utility in vasospasm prediction by evaluating several different ML networks and found that light gradient boosting machine (LightGBM) yielded the highest predictive power. Because our model uses an open source ML network and relies only on raw ICU data rather than human-generated clinical scoring systems or radiography, it is fully automated, deployable, and updatable as hospital admissions progress. Our model stratifies patients by CVRV risk level, allowing for customized hospital resource allocation for monitoring CV in each patient based on that individual's predicted risk.

Implications of all the available evidence

ML trained on a vast collection of routinely monitored ICU data yields a highly accurate and early predictor of CVRV. With a prospective study to further validate these findings, we believe that ML models can guide the development of individualized vasospasm monitoring plans based on predicted risk level as opposed to indiscriminate, resource-intensive monitoring in all patients, as is standard practice.

patients post-SAH to an ICU or neurocritical care unit (NCCU) for up to several weeks of vigilant, resource-intensive clinical monitoring.⁵

The ability to stratify SAH by CV risk would allow for clinical monitoring commensurate to that risk, as opposed to indiscriminate monitoring of every patient with the same vigilance. Accurately predicting CVRV would allow for the redirection of clinical resources towards higher-risk cases while limiting ICU resources to those unlikely to have severe symptoms and negative outcomes. There have been many research attempts to predict or draw associations to CV, most using either human-generated radiographic scores, clinical scores, or molecular biomarkers as predictors.^{6–9} Few studies have assessed the feasibility of predicting CV using ML,^{10–17} most achieving insufficient predictive accuracy for clinical rule-in/out of CV. Among the few studies achieving higher predictive accuracy, all were validated over small cohorts, were not externally validated, used a small portion of available clinical predictors, and relied on human-generated clinical scoring systems and subjectively interpreted radiographic findings, which are labor intensive and not available for all patients.^{10,14,17}

Our group was motivated to rigorously explore the utility of ML for vasospasm prediction in a large multicenter study. We tested the predictive accuracy of numerous machine learning networks trained on 172

unique ICU datapoints. We demonstrate the ability to predict CVRV with high accuracy using ML over the largest known cohort in the literature and over one week before the event occurs, and externally validate this model's utility at a separate institution.

Methods

Data extraction and inclusion criteria

Data was extracted from the Perioperative Data Warehouse, which was developed by the UCLA Department of Anesthesiology and Perioperative Medicine and distributed to multiple centers across the United States.¹⁸ All patients with an ICD-10 code of I60 (non-traumatic SAH) or S06.6 (traumatic SAH) between 2013 and 2022 at UCLA were included.

De-identified clinical data spanning the entire hospital admission was extracted for each patient. This information included basic demographics, vitals, routinely collected clinical labs (complete blood count with differential, basic metabolic panel, arterial blood gas, Hemoglobin A1c (HgA1c), and cerebrospinal (CSF) analysis), intracranial pressure (ICP), respiratory variables (O2 flow, FiO2, EtCO2, airway grade, intubation attempts, nitric oxide), fluid status variables (volume of maintenance intravenous (IV) fluid, urine output, blood loss, blood administered), feeding (gastric feeding,

emesis), and saturation (pulse oximetry, cerebral saturation)). Finally, the verapamil administration time was recorded if verapamil was administered. A complete list of clinical predictor variables is provided in [Supplemental Table S1](#).

Feature extraction

For all patients, clinical data collected before ICU admission and after verapamil injection time or time-of-discharge from the ICU (depending on if verapamil was given) were excluded. ICP and mean arterial pressure (MAPs) time-series were encoded into a 20-dimensional feature vector containing values corresponding to the 5th, 10th, ..., 95th, and 100th percentiles over the examined time period. Sera values such as complete blood count (CBC) results were encoded as a 5-dimensional feature vector representing the minimum, maximum, median, mean, and count of measurements. Additional “dependent variables” were derived from the originally extracted clinical variables, such as “total count of ICP measurements”.

Predictive model architecture, training, and cross validation

We developed two cross-sectional prediction models to aid in the evaluation of CVRV risk: a “prospective” predictive model and “retrospective” predictive model. The prospective model was developed to provide a clinically useful tool to assess CVRV risk in ICU patients. Prospective model predictions were performed using 4 h, 1 day, 3 days, 5 days, 7 days, and 10 days of ICU data starting from the time of ICU admission. Patients having yet to reach a primary endpoint were included at each prediction timepoint, resulting in progressively smaller prediction cohorts at each timepoint. A “binary” prospective model was trained to predict between two outcomes, namely if the patient would need verapamil or not. [Supplementary Figure S1](#) graphically illustrates the binary prospective model design. A more advanced “trinary” prospective model was trained to predict amongst three outcomes, namely “will never get verapamil”, “will get verapamil within three days”, or “will get verapamil after three days” from the time of prediction.

Retrospective models were developed to explore the events preceding CVRV by analyzing variable importance scores in groups of patients at the same chronological stage prior to vasospasm. For the retrospective model, predictions were performed using all ICU data starting from the primary endpoint and moving 4 h, 1 day, 3 days, 5 days, 7 days, and 10 days backwards in time, before the prediction target. [Supplementary Figure S1](#) graphically illustrates the retrospective model design.

Each model was tested with two different predictor sets: an “institutional” and “conservative” ([Supplemental Table S1](#)). The institutional predictor set contained all

clinical variables, whereas the conservative predictor set contained only variables which are strictly measured in a highly standardized manner across medical institutions. Examples of such standardized variables include vital signs, routine lab values, ICP, etc.

The candidate network architectures tested were Logistic Regression, K Nearest Neighbors, Naïve Bayes, Decision Trees, Support Vector Machines, Gaussian Process Classifiers, Ridge Classifier, Random Forest Classifier, Quadratic Discriminant Analysis, Ada Boost Classifier, Gradient Boosting Classifier, Linear Extra Trees Classifier, Extreme Gradient Boosting, Light Gradient Boosting Machine (LightGBM), and CatBoost Classifier. Logistic regression was chosen as a baseline model to compare against since prior medical literature has typically used logistic regression for linear predictive tasks (and we sought to determine if a non-linear machine learning approach was beneficial in comparison).¹⁹

Models were trained and tuned using the PyCaret²⁰ ML library (v3.1.0). Prior to training, we imputed missing features with average imputation. Features were normalized using Z score normalization as implemented in PyCaret. We employed a stratified five-fold cross-validation scheme (stratified by age, gender, and outcome) to report the average performance of models when trained on different subsets of the entire dataset. Stratification was done to ensure that there was equivalent representation of patient and clinical characteristics in both the training and testing sets.

To correct class imbalance in the training set, Synthetic Minority Over-Sampling Technique (SMOTE)²¹ was utilized with validation test distribution of values kept as-is. Model hyperparameters were further tuned using grid search (using the default PyCaret optimize function) on each model type to yield a higher AUC that was calibrated to the outcome across probability scores. The best model (highest AUC on the validation fold) was then set as the final model. The best model was found to be a LightGBM²² model in all analyses. We reported ROC curves and AUC values for each fold and calculated the average AUC (± 1 standard deviation (SD)) where 1 = best classifier, 0.5 = random classifier. We also created a PR (precision-recall) curve for each time point with an AP (average precision) value (± 1 SD). Lastly, we reported “Variable Importance Scores” (relative rankings of weights placed on factors used to train the model) for the retrospective models, to help interpret what characteristics the models focused on at different prediction time points. ROC curves were compared for statistical difference using the DeLong test. Code used to process the data and train the network is available at <https://github.com/abhisuri97/Vasospasm>.

External validation

Once models were trained and evaluated at UCLA, the associated code was sent to VUMC for local model

training and testing. We chose to train the model at VUMC separately since there were differences in data collection between the two institutions. Additionally, data harmonization between the two institutions was not possible due to the developers not being able to see output data from VUMC due to patient privacy concerns.

To train and test the model, VUMC data sets were created and formatted to match the inputs of the UCLA model in terms of general data types. Shared SQL queries ensured that the same data were extracted from the shared electronic health record vendor. Extracted clinical predictors were then manually mapped to ensure the raw data was of the expected values and format of the UCLA and inclusive of all similar VUMC data elements. There were several iterations of checking the dataset to verify VUMC applied the same patient inclusion criteria and procured the same set of clinical predictors. Once the data set was finalized, python scripts were executed locally to prepare the data for the model calibration and execution. All patient-level VUMC data remained on internal servers and only aggregated summary statistics were returned to the UCLA research team. Only the conservative set of inputs (as opposed to institutional set) was prepared for the VUMC model to simplify cross-center predictor variable mapping and limit predictions to those based only on highly standardized and widely available predictor variables. The three-group (trinary) prospective model was selected for external validation at VUMC as it represents the most difficult prediction scenario with three possible prediction outcomes as opposed to two.

Statistical methods

For this retrospective study, we aimed to include a sample size of all patients who we had any information on in our data warehouse to maximize applicability of predictive performance to a real life patient cohort. In terms of statistical tests, multiple statistical tests were undertaken in the course of this study. To determine if there were any differences in the amount of data collected per patient on a yearly basis, an ANOVA test was performed on the average number of datapoints collected for each patient in our dataset. ROC curves between institutional and conservative models were compared using the DeLong test. Significance level for all tests was set at $P = 0.05$ [DeLong].

Role of funders

Neither this study nor any other authors personally received financial support for the research presented in this manuscript, including study design, data collection, data analysis, interpretation, and writing of this report.

Ethics

This study qualified for UCLA IRB exception status (waiver of consent) because there was no direct contact with patients and all data in this study was de-identified.

Results

Demographics and sensitivity analysis

Demographics and case counts are summarized in [Table 1](#). A total of 1750 UCLA patients were included with an average age of 56 ± 20 years and 46% female. Verapamil was administered in 125 (7.1%) patients on average 7.6 ± 4.6 days after admission. A total of 1654 VUMC patients were included with an average age of 53 ± 21 years, 42% female, 75 (4.5%) receiving verapamil. Patient age distributions at each institution were similar, with most patients being over 50 years of age (63% at UCLA, 59% at VUMC). Racial distributions differed most significantly in the portion of Caucasian individuals, being 41% at UCLA and 84% at VUMC. The distributions of body mass index (BMI) at each institution did not differ significantly.

Furthermore, in order to assess whether there were any changes in the amount of data gathered on each patient by year, we performed an ANOVA test to determine if the average number of data points (sum of number of labs, ICP, and mean arterial pressure readings) for each patient differed by year. We found that there was no significant difference in the average number of datapoints collected on each patient by year ($P = 0.16$ [ANOVA]).

Prospective predictive models

In [Fig. 1](#), average ROC and PR curves for prospective UCLA models are reported. The binary prospective model achieved AUCs ranging from 0.68 (at $t = 10$ days of ICU data) to 0.88 (with $t = 4$ h of ICU data). The institutional and conservative model performances differed by 0.00–0.03 AUC points on average, which were not statistically significant differences ($P > 0.05$ for all [DeLong]). The trinary prospective model achieved AUCs in the range of 0.68–0.88 for predicting no CVRV, 0.73–0.83 for predicting CVRV <3 days, and 0.50–0.88 for predicting CVRV in 3+ days across the 1-, 5-, and 10-day timepoints. In [Fig. 2](#), LightGBM models are shown to outperform logistic regression at all timepoints. See [Supplementary Table S2](#) for tabularized prospective performance values. Additionally, theoretical CVRV rule-out performance is compared between models by examining precision when recall equals 1.00, with LightGBM outperforming logistic regression at all timepoints.

Retrospective predictive models

In [Fig. 3](#), average ROC and PR curves for binary retrospective UCLA models are reported. Average AUCs ranged from 0.81 to 0.92. The institutional and conservative model performances differed by 0.01–0.09 AUC points on average, which were not statistically significant differences ($P > 0.05$ for all [DeLong]). In [Fig. 4](#), AUCs of the binary retrospective models are compared to those of a control retrospective logistic regression model (see [Supplementary Table S3](#) for tabularized

Demographics	UCLA				VUMC			
	All	No Verapamil	Verapamil	p-value	All	No Verapamil	Verapamil	p-value
Count	n=1,750	n=1,625	n=125		n=1,654	n=1,579	n=75	-
Age								
Mean (SD)	56 (20)	56 (20)	51 (16)	0.005*	53 (21)	53 (21)	53 (15)	0.77
Median (Q1-Q3)	58 (43-71)	59 (43-71)	52 (41-63)	-	53 (34-70)	52 (35-70)	57 (44-65)	-
Age Category								
<18 years	55 (3%)	53 (3%)	2 (2%)	0.3	85 (5%)	85 (5%)	0	0.04*
18 to <50 years	517 (30%)	468 (29%)	49 (39%)	0.01*	568 (34%)	542 (34%)	26 (35%)	0.95
50 to <65 years	486 (28%)	439 (27%)	47 (38%)	0.01*	416 (25%)	386 (24%)	30 (40%)	0.002*
>= 65 years	620 (35%)	593 (36%)	27 (22%)	<0.001*	560 (34%)	541 (34%)	19 (25%)	0.11
Unkown	72 (4%)	72 (4%)	0	-	25 (2%)	25 (2%)	0	-
Sex								
F	799 (46%)	723 (44%)	76 (61%)	<0.001*	690 (42%)	642 (41%)	48 (64%)	<0.001*
M	951 (54%)	902 (56%)	49 (39%)	<0.001*	964 (58%)	937 (59%)	27 (36%)	<0.001*
Race								
Non-Hispanic White	724 (41%)	682 (42%)	42 (34%)	0.07	1,320 (84%)	1,264 (80%)	56 (75%)	0.26
Hispanic	437 (25%)	402 (25%)	35 (28%)	0.42	14 (1%)	12 (1%)	2 (3%)	0.08
Asian Pacific Islander	211 (12%)	194 (12%)	17 (14%)	0.58	4 (<1%)	4 (<1%)	0	0.66
Nonmspanic Black	176 (10%)	160 (10%)	16 (13%)	0.29	196 (12%)	185 (12%)	11 (15%)	0.44
American Indian/Alaska Native	7 (<1%)	7 (<1%)	0	0.46	0	0	0	n/a
Unknown/Other	195 (11%)	180 (11%)	15 (12%)	0.75	120 (7%)	114 (7%)	6 (8%)	-
BMI								
Mean (SD)	26 (6)	25 (7)	27 (5)	0.04*	27 (9)	27 (9)	29 (8)	0.078
Median (Q1-Q3)	25 (22-29)	25 (22-28)	26 (23-30)	-	26 (21-28)	26 (21-28)	27 (22-34)	-
BMI Category								
<18.5	97 (6%)	93 (6%)	4 (3%)	0.23	91 (6%)	88 (6%)	3 (4%)	0.56
18.5 to <25	686 (39%)	641 (39%)	45 (36%)	0.45	417 (25%)	395 (25%)	22 (29%)	0.4
25 to <30	536 (31%)	493 (30%)	43 (34%)	0.34	362 (22%)	345 (22%)	17 (23%)	0.87
>=30	310 (18%)	283 (17%)	27 (22%)	0.24	348 (21%)	322 (20%)	26 (35%)	0.003*
Unknown	121 (7%)	115 (7%)	6 (5%)	-	436 (26%)	429 (27%)	7 (9%)	-

An "*" indicates a statistically significant difference as defined by P < 0.05 [two-sample t-test for continuous measures, chi-squared for categorical measures].

Table 1: Basic demographics for the entire study cohort, the subset of patients not receiving verapamil, and the subset of patients receiving verapamil at each institution.

retrospective performance values). Additionally, the predictor variables ranked by average importance score (for LightGBM this is the number of times a feature was used to split data in each of the decision trees it creates), and the temporal fluctuation of the top three predictor variables preceding vasospasm are shown.

External validation

In Fig. 5, average ROC curves for the trinary prospective conservative VUMC models are reported. These models achieved AUCs ranging from 0.61 to 0.93 for predicting no CVRV, 0.81–0.82 for predicting CVRV <3 days, and 0.48–0.88 for predicting CVRV in 3+ days across the 1-, 5-, and 10-day timepoints. Predictive models at VUMC performed very similarly to those at UCLA (on average 0.01 AUC points lower).

Discussion

We describe our experience using ML to predict CV requiring verapamil in patients with SAH using a diverse and temporally granular collection of ICU data at two separate institutions. Predictive models at both institutions achieved comparable and high predictive

accuracy at multiple prediction time points, despite the patient populations differing appreciably in racial composition and being treated in separate health care systems with different decision pathways for verapamil administration.

Literature characterizing the utility of artificial intelligence (AI) and ML for vasospasm prediction is limited. One systematic review from 2021 on AI use in neurocritical care²³ identified a single study using sparse dictionary learning and covariance-based features from digital subtraction angiography (DSA) to predict vasospasm with an AUC of 0.93, but this study was limited by a sample size of n = 22 and DSA availability.¹⁷ Another 2021 systematic review of ML for stroke diagnosis and outcome prediction²⁴ reported one study predicting DCI with an AUC of 0.74 across n = 317 patients,¹² and one predicting a combination of DCI, angiographic vasospasm, or cerebral infarction with an accuracy of 95.2% but relied on radiography and matricellular protein lab availability.¹³

Among other studies employing AI to predict CV, Dumont et al. developed an artificial neural network (ANN) for predicting symptomatic cerebral vasospasm based on pre-existing clinical scores (Modified Fischer,

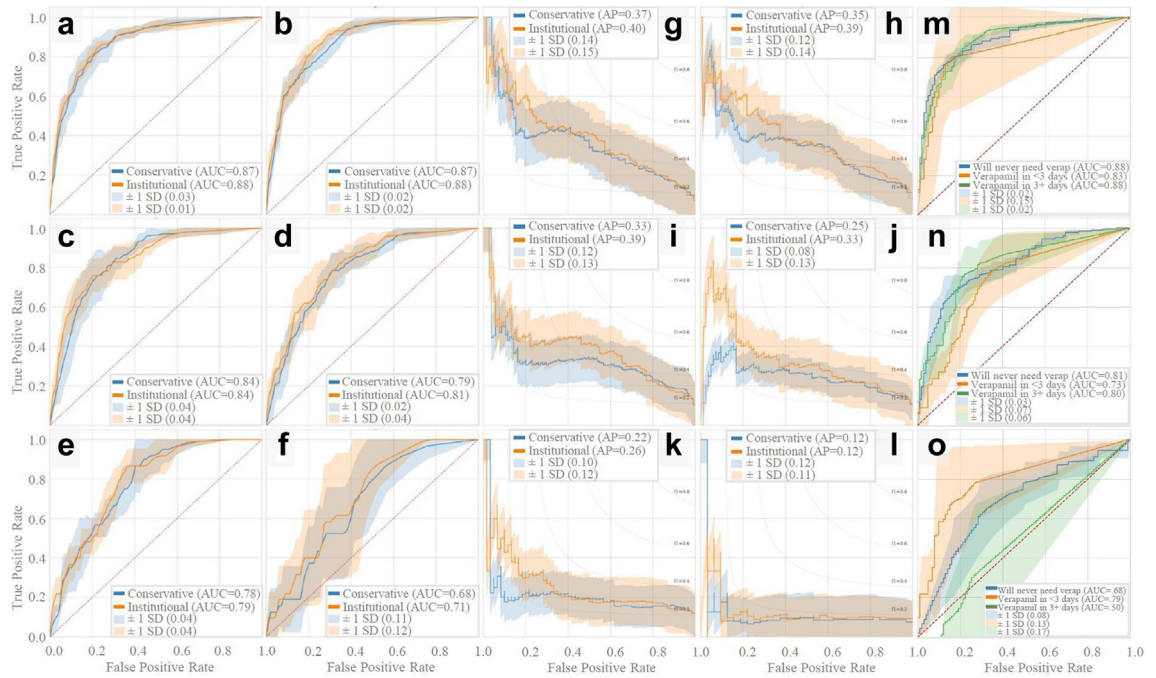


Fig. 1: Prospective model ROC and PR curves. Prospective model ROC and PR curves using five-fold cross validation. Binary prospective model results predicting CVRV or not during admission are shown in a–f. Panels a–f and g–l are AUCs and PR curves, respectively, using 4 h, 1 day, 3 days, 5 days, 7 days, and 10 days of ICU data since ICU admission. Trinary prospective model results predicting CVRV in ≤ 3 days, CVRV in >3 days, or no CVRV during admission, are shown in panels m–o, which show AUC curves using 1 day, 5 days, and 10 days of ICU data since ICU admission.

Hunt and Hess).¹⁴ The authors reported an AUC of 0.96, however this was only validated across $n = 22$ and relied on human-generated vasospasm risk scores. Skoch et al. adapted Dumont’s ANN for pediatric populations and achieved similar performance over $n = 16$ patients.¹⁵ Roederer et al. reported an AUC of 0.71 using a Naïve Bayes model on 81 patients with acute SAH within two days of vasospasm.¹⁶ Kim et al. predicted vasospasm defined as vessel stenosis $>50\%$ with an AUC of 0.88

($n = 343$) using a random forest¹⁰ reliant on clinical grading scales and manually extracted image features. We found that average AUCs for cohorts under $n = 1000$ were associated with significant uncertainty, justifying the need for larger cohorts.

Our prospective CVRV prediction model trained on raw ICU data alone achieved a high predictive accuracy (AUC = 0.88) over one week prior to verapamil injection and was validated over 1750 patients, which is the

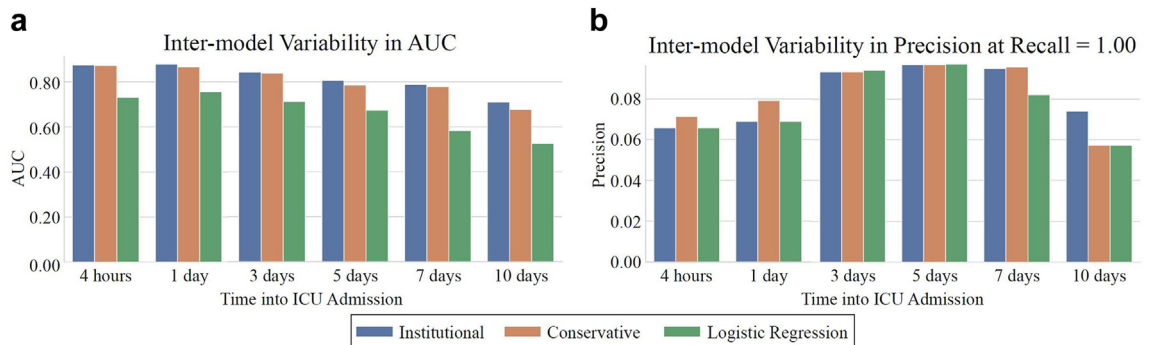


Fig. 2: Prospective model and logistic regression AUC and theoretical CVRV rule-out. Prospective institutional and conservative models and control model (logistic regression) AUCs and CVRV rule-out performance over time. Panel a shows AUCs at each prediction timepoint for each model. Panel b shows precisions at the threshold where Recall = 1.00.

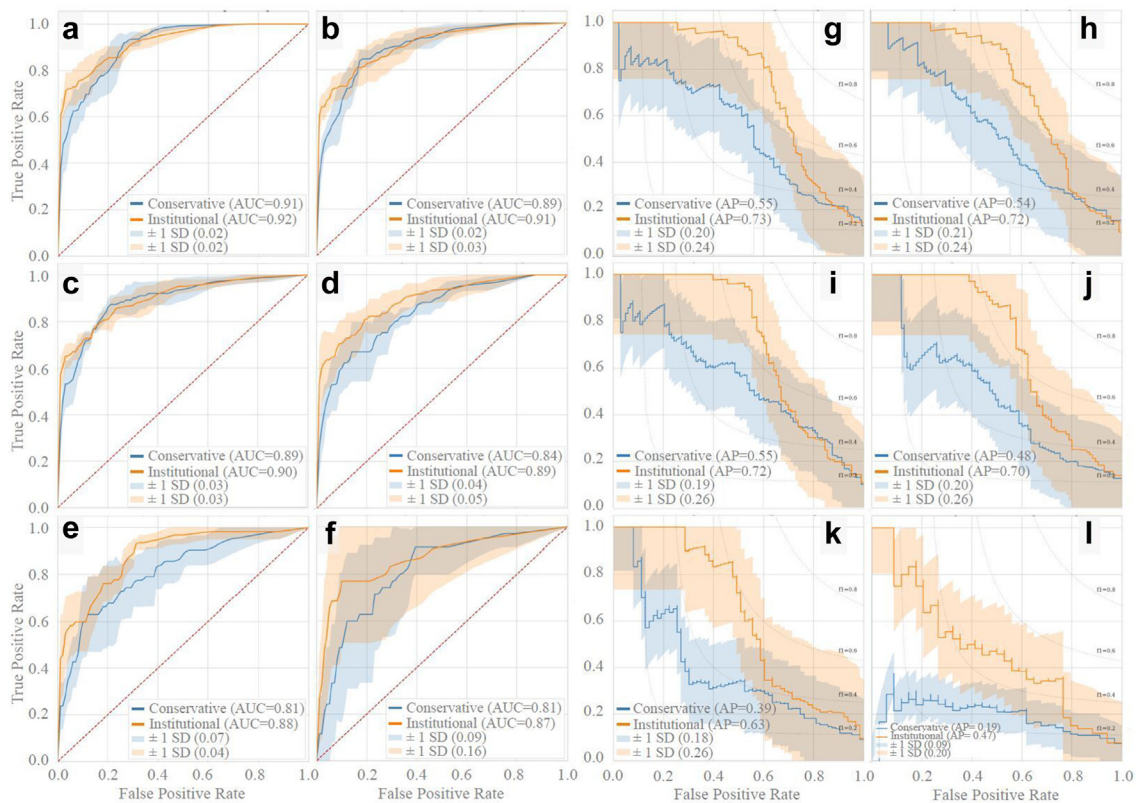


Fig. 3: Retrospective model ROC and PR curves. Retrospective model ROC and PR curves using five-fold cross validation. Binary retrospective model results predicting CVRV or not during admission are shown in panels a–l. Panels a–f and g–l are AUCs and PR curves, respectively, using 4 h, 1 day, 3 days, 5 days, 7 days, and 10 days of ICU prior to the primary endpoint (CVRV or no CVRV).

largest ML-driven vasospasm prediction cohort by a factor of five in the known literature, and the largest vasospasm prediction cohort of any prediction methodology in the literature. Our conservative model achieved a similar AUC of 0.87. Perhaps most notably, the PR curves for prospective models were favorable with regards to rule-out of patients who will not need verapamil, ruling out CVRV in 8% of those not requiring verapamil up to 10 days in advance and without any false negatives. Typical ICU management involves serial neuro checks \pm transcranial dopplers (TCDs) with subsequent computed tomography angiography (CTA) if suspicion for vasospasm is high enough, followed by endovascular intervention if warranted.⁵ Knowing a patient are unlikely to progress to CVRV is invaluable for a few reasons: 1) frequency of neuro checks and TCD can be decreased in this low-risk population, and 2) patients in this category who exhibit neurologic deficit with or without elevated TCDs can be managed medically (BP augmentation) rather than skipping directly to CTA for fear of CVRV, therefore potentially cutting down on CTA frequency in this population. Interestingly, predictive power worsened the

longer the length-of-stay in the ICU. As days pass in the ICU, the number of patients included in the prediction shrinks as patients have vasospasm and are therefore excluded from subsequent timepoints. It is possible that patient outcomes which are relatively easier to predict tend to present earlier in the admission with vasospasm, leaving a pool of patients whose primary endpoint is more difficult to discern. For example, by day 10, if a patient has not had vasospasm, it may become more difficult to predict that they will then have a vasospasm after day 10.

The trinary prospective model achieved high predictive accuracy (average AUC = 0.86) when predicting the timeline of impending vasospasm. Such predictions may serve to alert providers when to initiate prophylactic BP augmentation, for example, therefore enabling preparatory clinical action in high risk individuals and titration of monitoring vigilance based on an individual's CVRV risk. Because this is a prospective model like our previously described binary model, it can automatically generate new CVRV risk predictions every day for each patient in the ICU and guide care. Interestingly, as ICU admission progressed towards day 10,

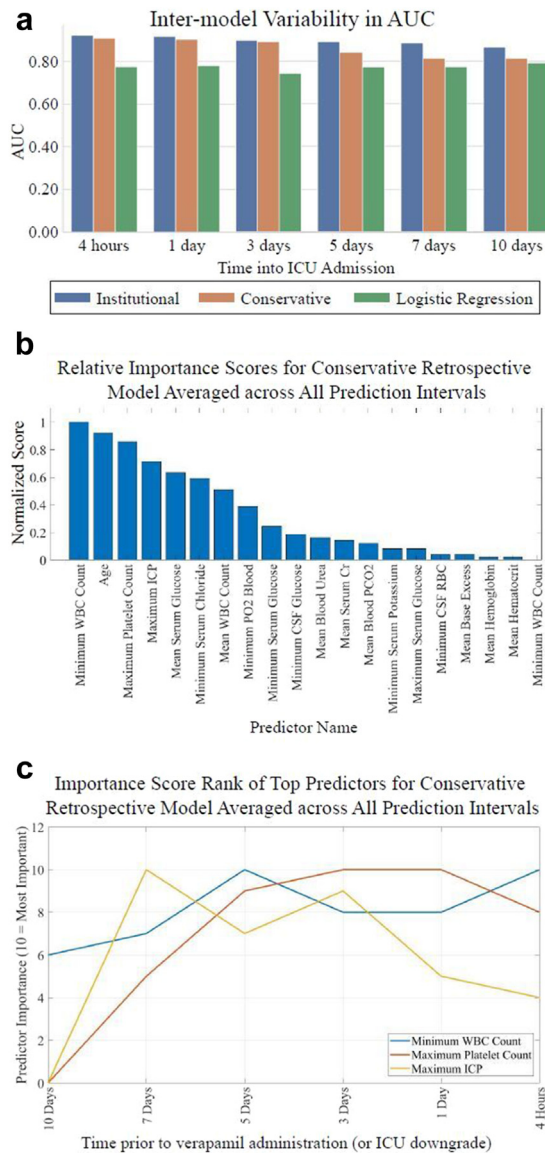


Fig. 4: Retrospective model and logistic regression AUCs and importance scores. Retrospective institutional, conservative, and control (logistic regression) models AUCs over time alongside retrospective conservative model importance score analysis. Panel a displays AUCs at each prediction timepoint for each model. Panels b and c show importance score analysis: b shows averaged ranked variable importance scores over all prediction timepoints and c shows temporal fluctuation in importance of the three predictor variables with the overall highest variable importance scores (10 = #1 overall predictor, 9 = #2 overall predictor, etc). Importance score rank of top three strongest predictor variables for the conservative model at each prediction interval.

predictive accuracy declined slightly for the binary prospective model and more significantly for the trinary prospective model. We suspect these declines are attributable to the steadily decreasing portion of patients

who are “obvious impending vasospasms” in the network’s eye.

By using only raw clinical data, our model inputs require zero human interpretation and can be fed clinical information without human supervision and update predictions in real-time. Prior attempts to predict vasospasm can be stratified into those which use only raw clinical data, and those which rely on clinical scores generated through human interpretation of diagnostic radiography. The former group has struggled to achieve predictive accuracy sufficient for clinical use.^{7,25} The other class of predictors^{9,10,13,26,27} rely on human-generated clinical scores which are time-intensive and limited by subjectivity inherent to human interpretation of radiographic features. Such predictors are also not easily updated in real-time given the practical limitations on imaging frequency in the clinical setting. Despite the convenience of our model’s fully automated workflow, its predictions must be interpreted by clinicians in the broader context of each patient’s admission, as blind acceptance of diagnostic medical AI tools may lead to medical errors and associated liability for physicians.²⁸ Additionally, our model is not proprietary and therefore readily implementable at any institution given all code is open source.

Our external validation supports the notion that our prospective models can be used in a multi-center context with nearly zero decline in peak predictive accuracy or CVRV rule-out performance. A logical next step will be to conduct a multi-center clinical study where it is prospectively predicted whether each patient will develop CVRV using our conservative model. Both the institutional and conservative models dramatically outperformed the control network, logistic regression, which assumes linearity between predictors and targets and is therefore performance limited in many settings. We therefore demonstrate a nonlinear relationship between the vast collection of ICU datapoints and the target.

Our study uniquely performed predictive modeling at multiple timepoints prior to the event. For the prospective model, doing so lends clinical utility in the form of daily risk CVRV predictions. With our retrospective model, performing predictions at multiple time points provides insight into the events leading to CVRV. Predictive accuracy began to decline appreciably in our retrospective model when predicting on a group of patients who were at least one week out from vasospasm. We believe the time interval prior to vasospasm at which retrospective predictive accuracy increases may represent the first detectable events in a pathophysiologic cascade towards vasospasm. To investigate this idea further, we analyzed the temporal fluctuation of predictor importance scores preceding verapamil in our retrospective conservative model.

Our analysis of the temporal fluctuation of vasospasm predictor importance scores may support

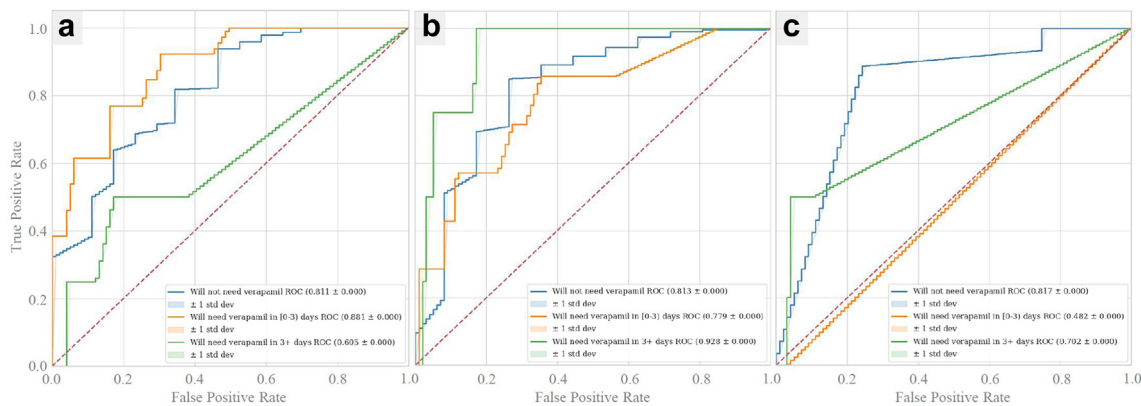


Fig. 5: External validation of vasospasm prediction. Prospective, trinary model trained and tested on the conservative set of clinical predictor variables from the VUMC dataset. Panel a displays ROCs and AUCs of five-fold cross validations for each group with the model trained on 1 day of ICU data, panel b with 5 days of ICU data, and panel c with 10 days of ICU data.

previously hypothesized CV pathophysiology. After initial intracranial haemorrhage and ICP elevation, subarachnoid blood products are thought to trigger microglial activation and macrophage “crosstalk” leading to peripheral immune activation.²⁹ Our model indicates that the predictive accuracy of maximum ICP peaks one week prior and then declines. This may explain the mixed results others have reported when attempting to predict CV with ICP alone within just a few days of onset.¹⁶ Our model shows that minimum white blood cell (WBC) count and maximum platelet count, markers of peripheral immune activation, rise in predictive accuracy approximately one week prior to vasospasm and immediately after maximum ICP importance score peaks. There is indeed considerable evidence that WBC and platelet counts are predictive of CV.^{30,31} The non-linear relationships between predictors and targets in ML do not allow us to infer that it is high, normal, low values, or some other characteristic entirely of any predictor which underlies its predictive accuracy. Therefore, the increases in the importance scores of minimum WBC and maximum platelet counts after maximum ICP predictive accuracy peaks may signal the initiation of the previously described peripheral immune activation following ICP elevation before CV.

While predictive models have been widely published across fields of medicine, they are rarely implemented in routine clinical practice. Many reasons may underlie the failure to implement, though lack of external validation is a well-known contributing factor. We would largely agree that this is a critical step towards clinician acceptance. The process of externally validating our UCLA predictive models proved to be complex. Despite using a common electronic health record vendor, mapping clinical predictor variables from UCLA to VUMC was a lengthy process with much back-and-forth between respective teams, chiefly because each institution was blinded to all patient-level data of the other. It was

necessary to have the VUMC clinical team verify clinical predictor variables were appropriately mapped, and a researcher with programming experience at VUMC to train and test the VUMC model. Given the anticipated complexity of the external validation process, we elected to test the most rigorous model configuration for simplicity, namely a model which prospectively predicted the previously described three-group outcome, using only the “conservative” set of clinical predictor variables. Our experience externally validating this model reinforces the notion that there is a large need in medicine to establish infrastructure to perform external validation quickly and easily.

Inclusion criteria were based on ICD codes for SAH and ICU time, which are only available after discharge; in a subsequent prospective analysis, another surrogate for inclusion will be required. Because ML is a “black box” which draws non-linear relationships between predictors and targets, we cannot know which specific characteristic of the most predictive variables spawn their predictive accuracy. It should be noted that there was moderate class imbalance in the cohorts within this study, having more non-verapamil patients than verapamil patients. This is a difficult problem to avoid in this setting, where the true number of those receiving verapamil is low relative to those not receiving verapamil. As previously described, best efforts were made to account for said class imbalance using SMOTE and reporting average precisions. Finally, this study is subject to limitations characteristic of retrospective analyses and hence requires prospective validation. Finally, both traumatic and atraumatic SAH was included in our prediction cohort; future work will involve building predictive models for specific etiologies of subarachnoid haemorrhage.

Cerebral vasospasm is a prevalent and life-threatening complication of SAH and requires vigilant clinical monitoring. Developing a reliable model for

predicting CV has been an area of ongoing research interest and has proven to be challenging. We report a highly accurate ML-driven predictor of CV requiring verapamil in, to our knowledge, the largest and only multi-center study in the literature. Our ML model analyzes 172 unique raw ICU datapoints to accurately predict CV requiring verapamil on average a full week in advance. Further research will focus on prospective validation of this model and the prediction of lesser forms of vasospasm to further optimize hospital resource allocation in this setting.

Contributors

DZ contributed to literature search, visualization study design, data collection, formal analysis, data interpretation, writing; AS contributed to data collection, formal analysis, visualization, writing; KM contributed to data collection, formal analysis, and writing; BG contributed to study design, formal analysis; BW contributed to study design, data collection, reviewing and editing; GC contributed to data interpretation, reviewing and editing; RF contributed to data interpretation, writing, reviewing and editing; EG contributed to visualization, study design, data collection, data analysis, data interpretation, and writing, and accessed and verified the underlying data reported in the manuscript and meets all four criteria for authorship outlined in the ICMJE Recommendations. All authors had full access to all the data in the study and accept responsibility to submit for publication. All authors read and approved of the final version of the manuscript.

Data sharing statement

The data utilized in this study was extracted from the Perioperative Data Warehouse¹⁸ and although de-identified, is not under the author's ownership and is therefore not available for public sharing.

Declaration of interests

The National Institutes of Health supports Vanderbilt University Medical Center, a participating institution in this study. Robert E. Freundlich is a consultant for Oak Hill Clinical Informatics, an expert witness for Hall Booth Smith, P.C., a Data and Safety Monitoring Board (DSMB) member for the Protocol and Statistical Analysis Plan for the Mode of Ventilation During Critical Illness (MODE) Trial, and the Treasurer for the Society of Technology in Anesthesia. Geoffrey Colby is a consultant for Medtronic, Stryker Neurovascular, and Rapid Medical. Eilon Gabel is a clinical trial consultant for Merck, Inc and unpaid co-founder of Extrico Health, Inc. All other authors have no conflict of interest. No author received financial support in conjunction with the generation of this submission.

Acknowledgements

We would like to acknowledge the late Dr. Bilwaj Gaonkar, co-author of this work and our dear friend. His lifetime contributions to the field of artificial intelligence are vast, and his joyful spirit will be forever remembered by those who knew him.

Appendix A. Supplementary data

Supplementary data related to this article can be found at <https://doi.org/10.1016/j.ebiom.2024.105206>.

References

- Dorsch NWC, King MT. A review of cerebral vasospasm in aneurysmal subarachnoid haemorrhage part I: incidence and effects. *J Clin Neurosci*. 1994;1(1):19–26 [cited 2022 Nov 17]. Available from: <https://pubmed.ncbi.nlm.nih.gov/18638721/>.
- EEG monitoring to detect vasospasm after subarachnoid hemorrhage [...]. [cited 2022 Nov 17]. Available from: <https://www.reliasmedia.com/articles/34004-eeeg-monitoring-to-detect-vasospasm-after-subarachnoid-hemorrhage>.
- Frontera JA, Fernandez A, Schmidt JM, et al. Defining vasospasm after subarachnoid hemorrhage. *Stroke*. 2009;40(6):1963–1968 [cited 2022 Nov 17]. Available from: <https://www.ahajournals.org/doi/abs/10.1161/STROKEAHA.108.544700>.
- Dabus G, Nogueira RG. Current options for the management of aneurysmal subarachnoid hemorrhage-induced cerebral vasospasm: a comprehensive review of the literature. *Interv Neurol*. 2013;2(1):30 [cited 2022 Nov 17]. Available from: <https://pubmed.ncbi.nlm.nih.gov/25957908/>.
- Diringer MN, Bleck TP, Hemphill JC, et al. Critical care management of patients following aneurysmal subarachnoid hemorrhage: recommendations from the Neurocritical care society's multidisciplinary consensus conference. *Neurocrit Care*. 2011;15(2):211–240 [cited 2022 Nov 17]. Available from: <https://pubmed.ncbi.nlm.nih.gov/21773873/>.
- Otite F, Mink S, Tan CO, et al. Impaired cerebral autoregulation is associated with vasospasm and delayed cerebral ischemia in subarachnoid hemorrhage. *Stroke*. 2014;45(3):677–682 [cited 2022 Nov 17]. Available from: <https://pubmed.ncbi.nlm.nih.gov/24425120/>.
- Przybycien-Szymanska MM, Ashley WW. Biomarker discovery in cerebral vasospasm after aneurysmal subarachnoid hemorrhage. *J Stroke Cerebrovasc Dis*. 2015;24(7):1453–1464 [cited 2022 Nov 17]. Available from: <https://pubmed.ncbi.nlm.nih.gov/25957908/>.
- Ishihara H, Oka F, Kawano R, et al. Hounsfield unit value of interpeduncular cistern hematomas can predict symptomatic vasospasm. *Stroke*. 2020;51(1):143–148 [cited 2022 Nov 17]. Available from: <https://www.ahajournals.org/doi/abs/10.1161/STROKEAHA.119.026962>.
- Frontera JA, Claassen J, Schmidt JM, et al. Prediction of symptomatic vasospasm after subarachnoid hemorrhage: the modified Fisher scale. *Neurosurgery*. 2006;59(1):21–26 [cited 2022 Nov 17]. Available from: <https://pubmed.ncbi.nlm.nih.gov/16823296/>.
- Kim KH, Koo HW, Lee BJ, Sohn MJ. Analysis of risk factors correlated with angiographic vasospasm in patients with aneurysmal subarachnoid hemorrhage using explainable predictive modeling. *J Clin Neurosci*. 2021;91:334–342 [cited 2022 Nov 17]. Available from: <https://pubmed.ncbi.nlm.nih.gov/34373049/>.
- Li J, Zhou K, Wang L, Cao Q. Predictive model of cerebral vasospasm in subarachnoid hemorrhage based on regression equation. *Scanning*. 2022;2022:3397967 [cited 2022 Nov 18]. Available from: <https://pubmed.ncbi.nlm.nih.gov/35581969/>.
- Ramos LA, van der Steen WE, Sales Barros R, et al. Machine learning improves prediction of delayed cerebral ischemia in patients with subarachnoid hemorrhage. *J Neurointerv Surg*. 2019;11(5):497–502 [cited 2022 Nov 17]. Available from: <https://pubmed.ncbi.nlm.nih.gov/30415227/>.
- Tanioka S, Ishida F, Nakano F, et al. Machine learning analysis of matricellular proteins and clinical variables for early prediction of delayed cerebral ischemia after aneurysmal subarachnoid hemorrhage. *Mol Neurobiol*. 2019;56(10):7128–7135 [cited 2022 Nov 17]. Available from: <https://pubmed.ncbi.nlm.nih.gov/30989629/>.
- Dumont TM, Rughani AI, Tranmer BI. Prediction of symptomatic cerebral vasospasm after aneurysmal subarachnoid hemorrhage with an artificial neural network: feasibility and comparison with logistic regression models. *World Neurosurg*. 2011;75(1):57–63 [cited 2022 Nov 17]. Available from: <https://pubmed.ncbi.nlm.nih.gov/21492664/>.
- Skoch J, Tahir R, Abruzzo T, Taylor JM, Zuccarello M, Vadivelu S. Predicting symptomatic cerebral vasospasm after aneurysmal subarachnoid hemorrhage with an artificial neural network in a pediatric population. *Childs Nerv Syst*. 2017;33(12):2153–2157 [cited 2022 Nov 17]. Available from: <https://pubmed.ncbi.nlm.nih.gov/28852853/>.
- Roederer A, Holmes JH, Smith MJ, Lee I, Park S. Prediction of significant vasospasm in aneurysmal subarachnoid hemorrhage using automated data. *Neurocrit Care*. 2014;21(3):444–450 [cited 2022 Nov 17]. Available from: <https://pubmed.ncbi.nlm.nih.gov/24715326/>.
- Capoglu S, Savarraj JP, Sheth SA, Choi HA, Giancardo L. Representation learning of 3D brain angiograms, an application for cerebral vasospasm prediction. *Annu Int Conf IEEE Eng Med Biol Soc*. 2019;2019:3394–3398 [cited 2022 Nov 17]. Available from: <https://pubmed.ncbi.nlm.nih.gov/31946608/>.
- Epstein RH, Hofer IS, Salari V, Gabel E. Successful implementation of a perioperative data warehouse using another hospital's published specification from epic's electronic health record system. *Anesth Analg*. 2021;132(2):465–474 [cited 2022 Nov 17]. Available from: <https://pubmed.ncbi.nlm.nih.gov/32332291/>.

- 19 Song X, Liu X, Liu F, Wang C. Comparison of machine learning and logistic regression models in predicting acute kidney injury: a systematic review and meta-analysis. *Int J Med Inform.* 2021;151:104484 [cited 2024 May 8]. Available from: <https://pubmed.ncbi.nlm.nih.gov/33991886/>.
- 20 Home - PyCaret [cited 2022 Nov 17]. Available from: <https://pycaret.org/>.
- 21 Chawla Nv, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. *J Artif Intell Res.* 2002;16:321–357 [cited 2022 Nov 17]. Available from: <https://www.jair.org/index.php/jair/article/view/10302>.
- 22 Ke G, Meng Q, Finley T, et al. LightGBM: a highly efficient gradient boosting decision tree [cited 2022 Nov 25]. Available from: <https://github.com/Microsoft/LightGBM>.
- 23 Azimi P, Mohammadi HR, Benzel EC, Shahzadi S, Azhari S, Montazeri A. Artificial neural networks in neurosurgery. *J Neurol Neurosurg Psychiatry.* 2015;86(3):251–256 [cited 2022 Nov 17]. Available from: <https://pubmed.ncbi.nlm.nih.gov/24987050/>.
- 24 Mainali S, Darsie ME, Smetana KS. Machine learning in action: stroke diagnosis and outcome prediction. *Front Neurol.* 2021;12:2153.
- 25 Zhang Y, Clark JF, Pyne-Geithman G, Caruso J. Metallomics study in CSF for putative biomarkers to predict cerebral vasospasm. *Metallomics.* 2010;2(9):628–637 [cited 2022 Nov 17]. Available from: <https://pubmed.ncbi.nlm.nih.gov/21072354/>.
- 26 Friedman JA, Goerss SJ, Meyer FB, et al. Volumetric quantification of fisher grade 3 aneurysmal subarachnoid hemorrhage: a novel method to predict symptomatic vasospasm on admission computerized tomography scans. *J Neurosurg.* 2002;97(2):401–407 [cited 2022 Nov 17]. Available from: <https://pubmed.ncbi.nlm.nih.gov/12186469/>.
- 27 Hickmann AK, Langner S, Kirsch M, et al. The value of perfusion computed tomography in predicting clinically relevant vasospasm in patients with aneurysmal subarachnoid hemorrhage. *Neurosurg Rev.* 2013;36(2):267–278 [cited 2022 Nov 17]. Available from: <https://pubmed.ncbi.nlm.nih.gov/23104502/>.
- 28 Price WN, Gerke S, Cohen IG. Potential liability for physicians using artificial intelligence. *JAMA.* 2019;322(18):1765–1766 [cited 2024 Apr 24]. Available from: <https://jamanetwork.com/journals/jama/fullarticle/2752750>.
- 29 Small C, Scott K, Smart D, et al. Microglia and post-subarachnoid hemorrhage vasospasm: review of emerging mechanisms and treatment modalities. *Clin Surg J.* 2022;3(3) [cited 2022 Nov 17]. Available from: <https://pubmed.ncbi.nlm.nih.gov/35450560/>.
- 30 McGirt MJ, Mavropoulos JC, McGirt LY, et al. Leukocytosis as an independent risk factor for cerebral vasospasm following aneurysmal subarachnoid hemorrhage. *J Neurosurg.* 2003;98(6):1222–1226 [cited 2022 Nov 17]. Available from: <https://pubmed.ncbi.nlm.nih.gov/12816268/>.
- 31 Aggarwal A, Salunke P, Singh H, et al. Vasospasm following aneurysmal subarachnoid hemorrhage: thrombocytopenia a marker. *J Neurosci Rural Pract.* 2013;4(3):257–261 [cited 2022 Nov 17]. Available from: <https://pubmed.ncbi.nlm.nih.gov/24250155/>.