ScientificScholar®
Knowledge is power
Publisher of Scientific Journals

www.surgicalneurologyint.com

Surgical Neurology International

Editor-in-Chief: Nancy E. Epstein, MD, Clinical Professor of Neurological Surgery, School of Medicine, State U. of NY at Stony Brook.

SNI: Trauma

Editor
Iype Cherian, MD
Krishna institute of Medical Sciences; Malkapur, India

SNI
Open Access

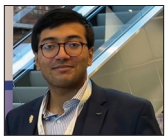
Original Article

Predicting neurosurgical referral outcomes in patients with chronic subdural hematomas using machine learning algorithms – A multi-center feasibility study

Sayan Biswas[1], Joshua Ian MacArthur[1], Anand Pandit[2], Lareyna McMenemy[1], Ved Sarkar[3], Helena Thompson[1], Mohammad Saleem Saleemi[4], Julian Chintzewen[1], Zahra Rose Almansoor[1], Xin Tian Chai[1], Emily Hardman[1], Christopher Torrie[1], Maya Holt[1], Thomas Hanna[1], Aleksandra Sobieraj[1], Ahmed Toma[5], K. Joshi George[4]

[1]Department of Medical Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester, [2]Department of High-Dimensional Neurology, University College London, London, United Kingdom, [3]Department of Computer Information Systems, De Anza College, Cupertino, United States, [4]Department of Neurosurgery, Manchester Centre for Clinical Neurosciences, Salford Royal Hospital, Manchester, United Kingdom, [5]Victor Horsley Department of Neurosurgery, National Hospital for Neurology and Neurosurgery, London, United Kingdom.

E-mail: *Sayan Biswas - sayan.biswas@srft.nhs.uk; Joshua Ian MacArthur - joshua.macarthur@nca.nhs.uk; Anand Pandit - a.pandit@ucl.ac.uk; Lareyna McMenemy - lareyna.mcmenemy@student.manchester.ac.uk; Ved Sarkar - sarkar.ved@gmail.com; Helena Thompson - helena.thompson-3@student.manchester.ac.uk; Mohammad Saleem Saleemi - mohammad.saleemi@nca.nhs.uk; Julian Chintzewen - julian.chintzewen@student.manchester.ac.uk; Zahra Rose Almansoor - zahra.almansoor@student.manchester.ac.uk; Xin Tian Chai - xintian.chai@student.manchester.ac.uk; Emily Hardman - emily.hardman-2@student.manchester.ac.uk; Christopher Torrie - christopher.torrie@student.manchester.ac.uk; Maya Holt - maya.holt@student.manchester.ac.uk; Thomas Hanna - thomas.hanna@student.manchester.ac.uk; Aleksandra Sobieraj - aleksandra.sobieraj@student.manchester.ac.uk; Ahmed Toma - a.toma@ucl.ac.uk; K. Joshi George - joshi.george@nca.nhs.uk



***Corresponding author:**
Sayan Biswas,
Faculty of Biology, Medicine and Health, University of Manchester, Manchester, United Kingdom.

sayan.biswas@srft.nhs.uk

Received : 04 December 2022
Accepted : 24 December 2022
Published : 20 January 2023

DOI
10.25259/SNI_1086_2022

Quick Response Code:

ABSTRACT

Background: Chronic subdural hematoma (CSDH) incidence and referral rates to neurosurgery are increasing. Accurate and automated evidence-based referral decision-support tools that can triage referrals are required. Our objective was to explore the feasibility of machine learning (ML) algorithms in predicting the outcome of a CSDH referral made to neurosurgery and to examine their reliability on external validation.

Methods: Multicenter retrospective case series conducted from 2015 to 2020, analyzing all CSDH patient referrals at two neurosurgical centers in the United Kingdom. 10 independent predictor variables were analyzed to predict the binary outcome of either accepting (for surgical treatment) or rejecting the CSDH referral with the aim of conservative management. 5 ML algorithms were developed and externally tested to determine the most reliable model for deployment.

Results: 1500 referrals in the internal cohort were analyzed, with 70% being rejected referrals. On a holdout set of 450 patients, the artificial neural network demonstrated an accuracy of 96.222% (94.444–97.778), an area under the receiver operating curve (AUC) of 0.951 (0.927–0.973) and a brier score loss of 0.037 (0.022–0.056). On a 1713 external validation patient cohort, the model demonstrated an AUC of 0.896 (0.878–0.912) and an accuracy of 92.294% (90.952–93.520). This model is publicly deployed: https://medmlanalytics.com/neural-analysis-model/.

Conclusion: ML models can accurately predict referral outcomes and can potentially be used in clinical practice as CSDH referral decision making support tools. The growing demand in healthcare, combined with increasing digitization of health records raises the opportunity for ML algorithms to be used for decision making in complex clinical scenarios.

Keywords: Artificial intelligence, Chronic subdural hematoma, Prediction tool, Referral decision making

INTRODUCTION

Chronic subdural hematoma (CSDH) is a common neurosurgical condition with a high prevalence in the elderly population.[31,38] Global incidence rates range from 1.72 to 20.6/100,000 persons/year,[5,14,23,29] a figure which is expected to increase with the aging population and the growing use of anti-coagulation/anti-platelet medications.[4,6,30,36] Typically, in the United Kingdom (UK), patients with CSDH present to outlying emergency departments and hospitals, where they are assessed by emergency physicians who complete the initial workup. Once a CT scan has identified a CSFH, irrespective of patient fitness or symptomatology, they are almost always referred to the on-call neurosurgical service at a tertiary center. These referrals generally have two outcomes: The acceptance or rejection of the referral by the neurosurgical team. The accepted patients are transferred for acute inpatient admission for considering surgical intervention and rejected referrals are advised for conservative medical management at the referring hospital, with re-referral in cases of neurological deterioration.

In the UK health system, patients with CSDH mostly present to the outlying district hospitals rather than the regional tertiary neurosurgery center and will be transferred across to the neurosurgery center only if there is a likelihood of needing surgical intervention. A large number of referrals are rejected for transfer as most patients are either not suitable for neurosurgery or do not need neurosurgery. These referrals, however, take up significant resources for both the referring physician as well as the neurosurgical team, who are already both overburdened with other clinical work. The development of an evidence-based decision tool for predicting CSDH referral outcomes that is easily interpretable by healthcare professionals can therefore be helpful. Such a tool would not only make a significant impact on the logistics of running an on-call neurosurgical service but also enable junior trainees on-call and emergency departments to be more confident in making the decision to refer as well as accept based on the tool.

Machine learning (ML) is a rapidly developing field that can be applied to health-related outcome prediction models. ML algorithms can identify complex non-linear relationships between a set of features and an output, generating interpretable results in real time.[33] There is a paucity of literature in the different techniques utilized to predict the outcome of CSDH referral, with the presence of only one previous study applying conventional statistical methods to the problem.[8] These models can suffer from over or under-training during the initial development process resulting in misleading algorithm performances, that are only evident upon external validation at different care centers with different patient demographics. The ML models will also be able to provide patient specific results on the influence of their clinical and radiological parameters on referral outcome, thereby supporting the clinical reasoning and patient centered decision-making ability of physicians in training.

The aim of this study was to create an ML model capable of replicating neurosurgical decision making and evaluate its reliability in predicting acceptance of CSDH referrals in two separate neurosurgical centers.

MATERIALS AND METHODS

Guidelines

The Transparent Reporting of Multivariable Prediction Models for Individual Prognosis or Diagnosis checklist and the JMIR Guidelines for Developing and Reporting ML Predictive Models in Biomedical Research were followed in our analysis.[12,25]

Data source and feature selection

The initial study was a single center retrospective analysis of all CSDH patient referrals to the local neurosurgery unit in Manchester from 2015 to 2020. The local trust's neurosurgical referral database was examined and all CSDH referrals were exported in an anonymized manner for analysis. The exclusion criteria included any patients with missing data and those without a decision given at the time of initial referral. 16 neurosurgery consultants and 18 registrars were involved in the decision making of these CSDH referrals. A total of 1500 referrals were identified, of which 450 referrals were accepted referrals and the remaining 1050 were rejected referrals. These 1500 patient referrals were used for the development and validation of the models. Patient consent was not required as the study was conducted in an anonymized and retrospective manner. The study was approved by the local hospital's research and innovation board, reference number: 22HIP11.

Ten predictor variables were identified on the referral database and the relevant information was extracted at the point of first referral: age (continuous), sex of the patient (male or female), GCS of the patient at referral (discrete continuous), presence of headache (yes or no), dementia (yes or no), motor weakness (yes or no), midline shift (yes or no), the size of the CSDH (small, medium or large; determined from radiological reports), the patient's pre-morbid quality of life (QoL) (poor or reasonable), and their anti-coagulation status (yes or no). The presence of midline shift on radiological scans was based on the referring neuro-radiology report and the hematoma sizes were defined using the maximal hematoma thickness; with hematomas <1 cm being considered small, 1–2 cm as medium and >2 cm as large. Poor pre-morbid QoL was defined as patients requiring full time care or those classified as an American Society of Anesthesiologists Grade 4, with all other patients classified as having a reasonable pre-morbid QoL. The binary outcome

variable of the model was the acceptance or rejection of a CSDH referral, with acceptance defined as admitting the patient to the neurosurgery center for assessment with the aim of proceeding to urgent surgical intervention. Predictor variable selection was conducted through stepwise multivariable logistic regression analysis and recursive feature elimination (RFE) using a stratified 4-fold cross validation with 4 repeats to determine the optimal number of variables employed in the ML models.

Model analysis

A stratified 70:30 train – test split was carried out on the total cohort of 1500 patient referrals, with 1050 data points utilized for training the models. All continuous data were centered to zero mean and scaled to unit variance while all categorical data were dummy coded. 6 ML models were created; 5 supervised learning algorithms: Logistic regression model, Support Vector Machine (SVM), K-nearest neighbors (KNN), Decision Tree (DT) and 1 deep learning multi-level perceptron artificial neural network (ANN) framework. Model hyperparameters were optimized through an iterative process that calibrated the weight estimations for each model based on their best yield for accuracy.

The models were trained on 4-fold stratified K-fold cross validation with 100 repeats on the training dataset. In stratified K-fold cross-validation, the class ratio for each fold is equivalent to the class ratio of the original dataset for each variable.[7] This allowed for standardizing the class imbalances present within our predictor variables and provided us with the best average performance results for the models.[21] The performance of the models was evaluated via 6 performance metrics on the training and testing sets: accuracy, recall/sensitivity, precision/positive predictive value, specificity, area under the receiver operating curve (AUC)/discrimination, and the brier score loss (refer to Supplementary Table 1 for definitions). All metrics were bootstrapped with 1000 resamples to derive the associated 95% confidence intervals (CI). The best performing model was selected after comparative analysis of the different model performance metrics on the training and testing sets: DeLong's test was used to compare the difference in the AUC of any two models on the testing set, the McNemar's test was used to compare the difference in the accuracy of two models on the testing set and the Mann–Whitney U test was used to compare the difference in the mean sensitivity of the 16 different training folds of any two models on the training set. Finally, decision curve analysis (DCA) was performed to confirm the model that provided the greatest clinical net benefit when predicting the outcome of a CSDH referral over a wide range of predicted threshold probabilities. The net benefit is defined as a function of the number of true positives, false positives, and the threshold probability. The DCA plot allows a physician to determine a threshold best suited for an individual patient's need and evaluate the net benefit of using the ML model relative to the default strategies of treating all or no patients.[40]

The best performing model was then calibrated on the testing test. Calibration is a measure of the consistency between the model's predicted probabilities and the true observed probabilities in the study population. It is depicted as a calibration curve, that is ideally a 45–° line through the origin, with a slope of 1 (an assessment of the spread of the estimated probabilities/risk compared to the observed

Supplementary Table 1: Description of the performance metrics used for evaluating the ML models in context of the study.

Performance metrics	Description
Accuracy	The percentage of correctly predicted outcomes compared to all actual true outcomes.[43]
Recall/sensitivity	Recall is the ability of our model to correctly identify the patients accepted into neurosurgery out of all the patients predicted by the model to be accepted into neurosurgery.[19]
Precision/positive predictive value	Precision is the ability of our model to correctly identify the patients accepted into neurosurgery out of all the patients originally accepted into neurosurgery.[19]
Specificity	Specificity is the ability of our model to correctly identify the patients rejected by neurosurgery out of all the patients predicted by the model to be rejected.[37]
AUC/Discrimination/ c-statistic	AUC is the numerical representation of the discrimination of a model. It is the ability of the model to distinguish between patient referrals that are accepted into neurosurgery compared to those that are not and the values range from 0.5 (no better than random chance) to 1 (perfect discrimination).[9]
Brier score loss	Measures the mean squared difference between the predicted probability and the actual outcome. Brier scores range from 0 to 1 and values close to 0 indicate better calibrated models with less error between the predicted and outcome outcomes.[15,34]
SHAP feature importance values	SHAP feature importance is feature importance ranked based on the Shapley values. Shapley values give the marginal contributions of a feature across all permutations of a model during cross-validations. This allows the analysis of a feature globally across all the models created during cross validation. Mean|SHAP|values then provide the absolute value of the average feature importance scores across all cross-validation folds.[26]

SHAP: SHapley additive exPlanation, ML: Machine learning, AUC = Area under the receiver operating curve

Table 1: Cohort demographics of 1500 patient referrals with descriptive statistical analysis using t-tests for continuous variables (Mean [± Standard Deviation]) and Chi-square tests for categorical variables (*n* [%]).

	Rejected cohort (*n*=1050) (%)	Accepted cohort (*n*=450) (%)	Total cohort (*n*=1500) (%)	*P*-value
Age	78.352±13.408	71.582±14.592	76.321±14.114	6.999×10^{-18}
Sex				
Male	733 (48.8)	346 (23.1)	1079 (71.9)	0.005
Female	317 (21.1)	104 (7.0)	421 (28.1)	
Headache				
Yes	135 (9.0)	167 (11.1)	302 (20.1)	6.981×10^{-27}
No	915 (61.0)	283 (18.9)	1198 (79.9)	
Dementia				
Yes	231 (15.4)	30 (2.0)	261 (17.4)	7.054×10^{-13}
No	819 (54.6)	420 (28.0)	1239 (82.6)	
GCS	14.170±1.620/15 (14–15)	14.289±1.610/15 (14–15)	14.205±1.613/15 (14–15)	0.189
Motor weakness				
Yes	110 (7.3)	177 (11.8)	287 (19.1)	9.353×10^{-39}
No	940 (62.7)	273 (18.2)	1213 (80.9)	
Midline shift				
Yes	89 (5.9)	337 (22.5)	426 (28.4)	1.311×10^{-150}
No	961 (64.1)	113 (7.5)	1074 (71.6)	
Size of CSDH				
Small	839 (56.0)	11 (0.7)	850 (56.7)	3.171×10^{-220}
Medium	141 (9.4)	42 (2.8)	183 (12.2)	
Large	70 (4.7)	397 (26.4)	467 (31.1)	
Pre-morbid QoL				
Poor	155 (10.3)	9 (0.6)	164 (10.9)	3.919×10^{-13}
Reasonable	895 (59.7)	441 (29.4)	1336 (89.1)	
Anti-coagulation				
Yes	453 (30.2)	165 (11.0)	618 (41.2)	0.020
No	597 (39.8)	285 (19.0)	882 (58.8)	

CSDH: Chronic subdural hematoma, GCS: Glasgow coma scale, QoL: Quality of life

probabilities) and an intercept of 0 (measure of the model's tendency of under-estimating (<0) or over-estimating (more than 0) the true probability of the dataset).[34,35,39] In this study, Platt scaling/sigmoid binned calibration spread across 10 bins was the preferred method of calibration, and the models were evaluated through analysis of the shape of the calibration belt, its slope, intercept, and the Brier score loss metric.

Model agnostic interpretation was conducted on the trained models through partial dependence plots (PDPs) and feature importance calculations. PDPs helped visualized the impact of a predictor variable on the outcome of the model by marginalizing over all the values of the input variable.[16] In addition, feature importance calculations for all predictor variables were performed on the optimal model using the SHapley Additive exPlanation (SHAP) method, to produce mean SHAP scores for each variable on a scale from 0 to 1 [Supplementary Table 1].

All ML analysis was performed using the R coding language version 3.4.3 (The R Foundation, Vienna, Austria), Python coding language (version 3.8, Python Software Foundation, Wilmington, Delaware) in an Anaconda virtual environment (Anaconda Inc., Austin, Texas), using TensorFlow 2.1 for deep learning,[1,2] and the Scikit-learn package for supervised learning.[11,28] IBM Statistical Package for the Social Sciences (SPSS) software (SPSS Inc., Chicago, IL, USA) Version 25 for Mac was utilized for all the statistical analysis. $P < 0.05$ was considered statistically significant. Step-wise multivariable logistic regression analysis was performed to determine the impact of each predictor variable on the outcome.

External validation

A retrospective analysis was conducted of all CSDH patients referred to a neurosurgical center in London, to encompass a larger, more heterogenous patient population, spanning 5 years from 2015 to 2020. A total of 2200 patient referrals were identified but only 1713 were selected for analysis following removal of duplicates and missing/inappropriate data. Of the 1713 referrals identified 505 were accepted referrals and 1204 were rejected referrals. These 1713 patient referrals were used for external validation of the optimal ML model.

Application deployment

The optimal ML model was incorporated into a natively developed interactive web application, that is, publicly deployed and is optimized for use on all major desktops, tablets, and mobile phones. This open access clinical tool will allow primary and secondary care health-care professionals from all over the world to input in values into the pre-trained model and retrieve an output in real time.

RESULTS

Baseline patient characteristics

The average age of the cohort was 76.321 ± 14.114 years and 71.9% of the patients were male. Detailed cohort demographic information for this study is summarized in Table 1. Majority of patients in the patient cohort had no headache (79.9%), dementia (82.6%), motor weakness (80.9%) nor midline

Table 2: Performance metrics of the 5 ML models on the testing set with 95% confidence intervals.

Model	Accuracy (%)	Recall/Sensitivity (%)	Precision/PPV (%)	Specificity (%)	AUC	Brier score loss
Testing Set (*n*=450)						
Logistic regression	94.667 (92.444–96.667)	92.908 (88.535–96.855)	90.345 (85.156–94.891)	95.469 (92.343–97.404)	0.942 (0.917–0.964)	0.053 (0.033–0.076)
SVM	93.556 (91.111–95.778)	90.780 (85.816–95.302)	88.889 (83.562–93.878)	94.822 (91.561–96.913)	0.928 (0.900–0.953)	0.064 (0.042–0.089)
KNN	94.222 (92.000–96.222)	90.780 (86.207–95.364)	90.780 (86.207–95.364)	95.729 (92.738–97.645)	0.933 (0.907–0.957)	0.058 (0.038–0.080)
DT	96.000 (94.000–97.778)	92.199 (87.324–96.241)	94.891 (90.909–98.374)	97.735 (95.186–99.004)	0.949 (0.924–0.972)	0.040 (0.022–0.060)
ANN	96.222 (94.444–97.778)	92.199 (87.500–96.250)	95.588 (91.667–98.582)	98.058 (95.613–99.208)	0.951 (0.927–0.973)	0.037 (0.022–0.056)

ANN: Artificial neural network, AUC: Area under the receiver operating curve, DT: Decision tree, KNN: K-nearest neighbours, PPV: Positive predictive value, SVM: Support vector machine, ML: Machine learning

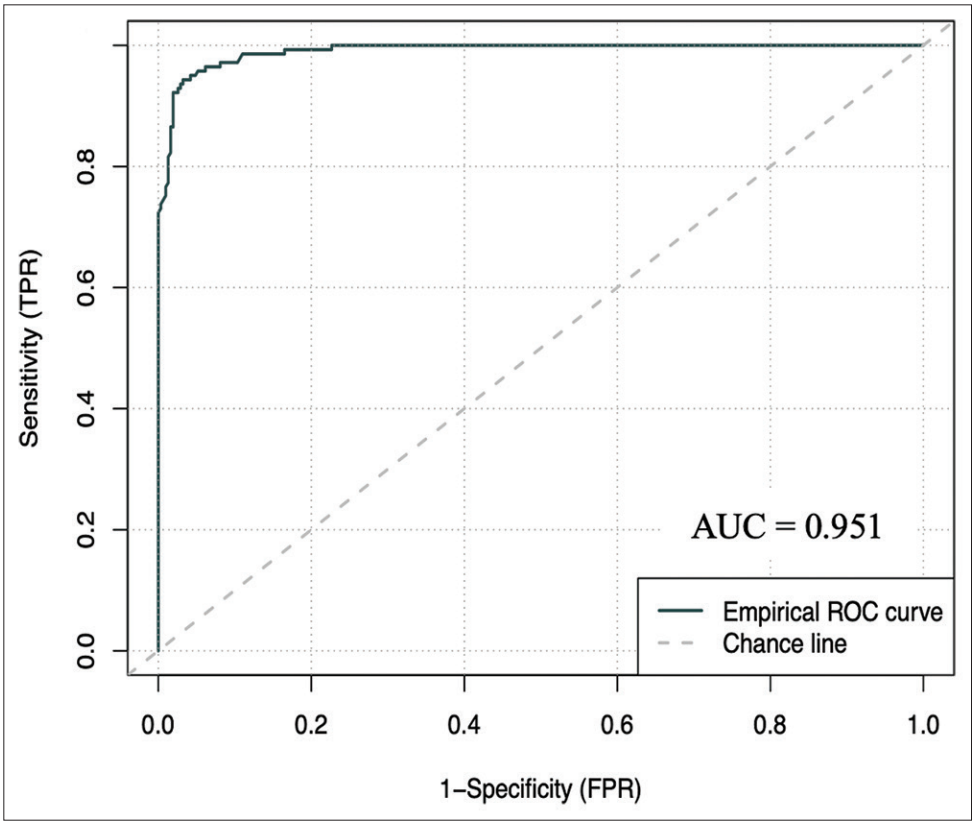


Figure 1: Receiver operating curve of the artificial neural network model for predicting acceptance of chronic subdural hematoma referrals, AUC = 0.951, on the testing set, *n* = 450. AUC = Area under the receiver operating curve.

shift (71.6%) at the time of the referral. In addition, most patients had a small sized hematoma (56.7%), a reasonable pre-morbid QoL (89.1%) and were not on anti-coagulation (58.8%). Independent samples test results revealed a statistically significant difference in the mean age of each outcome class, with the accepted cohort being significantly younger than the rejected cohort ($P = 6.999 \times 10^{-18}$). The accepted cohort also had a comparatively higher mean GCS, but there was no significant difference ($P = 0.189$). Further stepwise multivariable logistic regression analysis demonstrated that the age of the patient (OR: 0.998 [0.997–0.999], $P = 0.001007$), the presence of headaches (OR: 1.093 [1.056–1.132], $P = 4.17 \times 10^{-7}$), dementia (OR: 0.936 [0.903–0.970], $P = 0.000247$), motor weakness (OR: 1.108 [1.070–1.147], $P = 8.84 \times 10^{-9}$), midline shift (OR: 1.136 [1.086–1.188], $P = 2.83 \times 10^{-8}$), the size of the CSDH (OR: 1.389 [1.357–1.421], $P \leq 2.0 \times 10^{-16}$), and the pre-morbid QoL of the patient (OR: 0.841 [0.806–0.878], $P = 3.81 \times 10^{-15}$) were all statistically significant predictors of acceptance of a CSDH referral. Thus, only these seven variables were used as the predictor variables for the ML models.

Model performance

Supplementary Table 2 demonstrates the performance metrics for the 5 ML models on the training dataset. These results clearly indicate that the best trained model was the DT model with an accuracy of 98.762% (95% CI: 98.095–99.333), sensitivity of 97.411% (95% CI: 95.469–99.032), specificity of 99.325% (95% CI: 98.337–99.751), PPV of 98.366% (95% CI: 96.865–99.671), an AUC of 0.984 (95% CI: 0.974–0.992) and a brier score loss of 0.012 (95% CI: 0.007–0.019). Subsequently, the performance of all 5 ML models was evaluated on the testing set [Table 2]. These results suggested that the ANN model was the optimal model with an accuracy of 96.222% (95% CI: 94.444–97.778), sensitivity of 92.199% (95% CI: 87.500–96.250), specificity of 98.058% (95% CI: 95.613–99.208), PPV of 95.588% (95% CI: 91.667–98.582), an AUC of 0.951 (95% CI: 0.927–0.973) [Figure 1], and a brier score loss of 0.037 (95% CI: 0.022–0.056). Comparative statistical analysis between the performance metrics of the 5 ML models on the testing set demonstrated that the ANN was statistically significantly better than the SVM and the KNN models but not the DT and logistic regression models [Supplementary Table 3]. DCA then further highlighted that the ANN model provided greater clinical net benefit compared to the DT and logistic regression models at all predicted probabilities relative to default strategies of management for all or no patients [Figure 2]. Thus, the ANN was considered to be the most best performing model for predicting the outcome of a CSDH referral. Figure 3 illustrates the calibration curve of this ANN model on the testing set with an intercept of −0.44 (−0.95–

Table 3: Differences between the baseline characteristics of the internal cohort and external validation datasets.

	Internal cohort (*n*=1500)	External validation (*n*=1713)	*P*-value
Age	76.321±14.114	76.695±14.423	0.089
Headache			
Yes	302 (20.1%)	430 (25.1%)	0.485
No	1198 (79.9%)	1283 (74.9%)	
Dementia			
Yes	261 (17.4%)	243 (14.2%)	0.833
No	1239 (82.6%)	1470 (85.8%)	
Motor weakness			
Yes	287 (19.1%)	488 (29.5%)	0.171
No	1213 (80.9%)	1225 (71.5%)	
Midline shift			
Yes	426 (28.4%)	585 (34.2%)	0.745
No	1074 (71.6%)	1128 (65.8%)	
Size of CSDH			
Small	850 (56.7%)	788 (46%)	0.746
Medium	183 (12.2%)	421 (24.6%)	
Large	467 (31.1%)	504 (29.4%)	
Pre-morbid QoL			
Poor	164 (10.9%)	474 (27.7%)	0.415
Reasonable	1336 (89.1%)	1239 (72.3%)	
Acceptance status			
Accepted	450 (30%)	505 (29.5%)	0.747
Rejected	1050 (70%)	1208 (70.5%)	

Categorical variables represented as *n* (%), continuous variables represented as mean (± Standard Deviation). CSDH: Chronic subdural hematoma, QoL: Quality of life. **P* value significant at the 0.05 level

Supplementary Table 2: Performance metrics of the 5 ML models on the training set, *n*=1050, with 95% confidence intervals.

Model	Accuracy (%)	Recall/sensitivity (%)	Precision/PPV (%)	Specificity (%)	AUC	Brier score loss
Training Set (*n*=1050)						
Logistic regression	93.619 (92.191–95.048)	88.673 (85.374–92.151)	89.543 (86.007–93.016)	95.682 (93.889–96.981)	0.922 (0.904–0.9415)	0.064 (0.050–0.079)
SVM	91.619 (89.905–93.238)	87.792 (84.971–91.096)	84.424 (80.374–88.401)	93.252 (91.139–94.903)	0.905 (0.884–0.924)	0.084 (0.068–0.101)
KNN	94.286 (92.857–95.619)	90.291 (86.897–93.312)	90.291 (86.897–93.312)	95.951 (94.201–97.206)	0.931 (0.913–0.948)	0.057 (0.044–0.071)
DT	98.762 (98.095–99.333)	97.411 (95.469–99.032)	98.366 (96.865–99.671)	99.325 (98.337–99.751)	0.984 (0.974–0.992)	0.012 (0.007–0.019)
ANN	94.667 (93.333–96.000)	88.673 (85.209–92.079)	92.881 (89.892–95.623)	97.166 (95.626–98.193)	0.922 (0.912–0.948)	0.053 (0.040–0.067)

ANN: Artificial neural network, AUC: Area under the receiver operating curve, DT: Decision tree, KNN: K-nearest neighbours, PPV: Positive predictive value, SVM: Support vector machine, ML: Machine learning

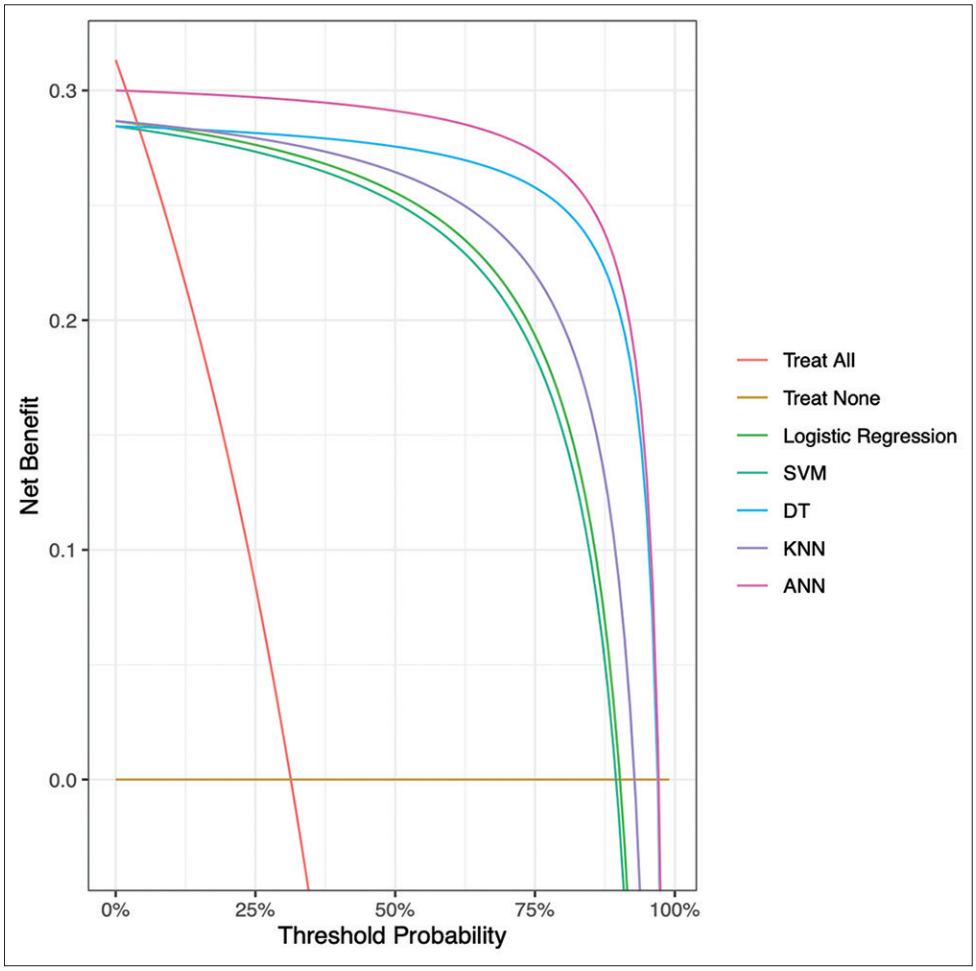


Figure 2: Decision curve analysis comparing expected clinical net benefit of the 5 different machine learning models on the testing set. SVM = Support Vector Machine, KNN = K-nearest neighbours, DT = Decision Tree, ANN = Artificial Neural Network, ANCHOR = Artificial Neural network for Chronic subdural HematOma Referral outcome prediction.

0.08) and a slope of 0.99 (0.78–1.20). This ANN model, named ANCHOR (an ANN for CSDH Referral outcome prediction) has been deployed as a web application at https://medmlanalytics.com/neural-analysis-model/.

Supplementary Table 3: Statistical significance of the comparison between the performance of all machine learning models on the training and testing set.

Model	Logistic regression	SVM	KNN	DT	ANN
Logistic regression	-	0.08261	0.6978	0.6539	0.304
		0.752	0.540	0.137	0.096
		0.06576	<0.00001*	<0.00001*	<0.00001*
SVM	0.08261	-	0.1093	0.2191	0.038*
	0.752		0.286	0.082	0.009*
	0.06576		0.00142*	<0.00001*	<0.00001*
KNN	0.6978	0.1093	-	0.8331	0.025*
	0.540	0.286		0.404	0.015*
	<0.00001*	0.00142*		<0.00001*	<0.00001*
DT	0.570	0.135	0.207	-	0.889
	0.137	0.082	0.404		0.990
	<0.00001*	<0.00001*	<0.00001*		<0.00001*
ANN	0.304	0.038*	0.025*	0.889	-
	0.096	0.009*	0.015*	0.990	
	<0.00001*	<0.00001*	<0.00001*	<0.00001*	

DLT: DeLong's test, MCN: McNemar's test, MWU: Mann-whitney u test for comparison of the AUC and Accuracy of the models on the testing set and recall of the models on the training set. *Statistically significant at $P<0.05$. SVM = Support vector machine, KNN = K-nearest neighbours, DT = Decision tree, ANN = Artificial neural network, AUC = Area under the receiver operating curve

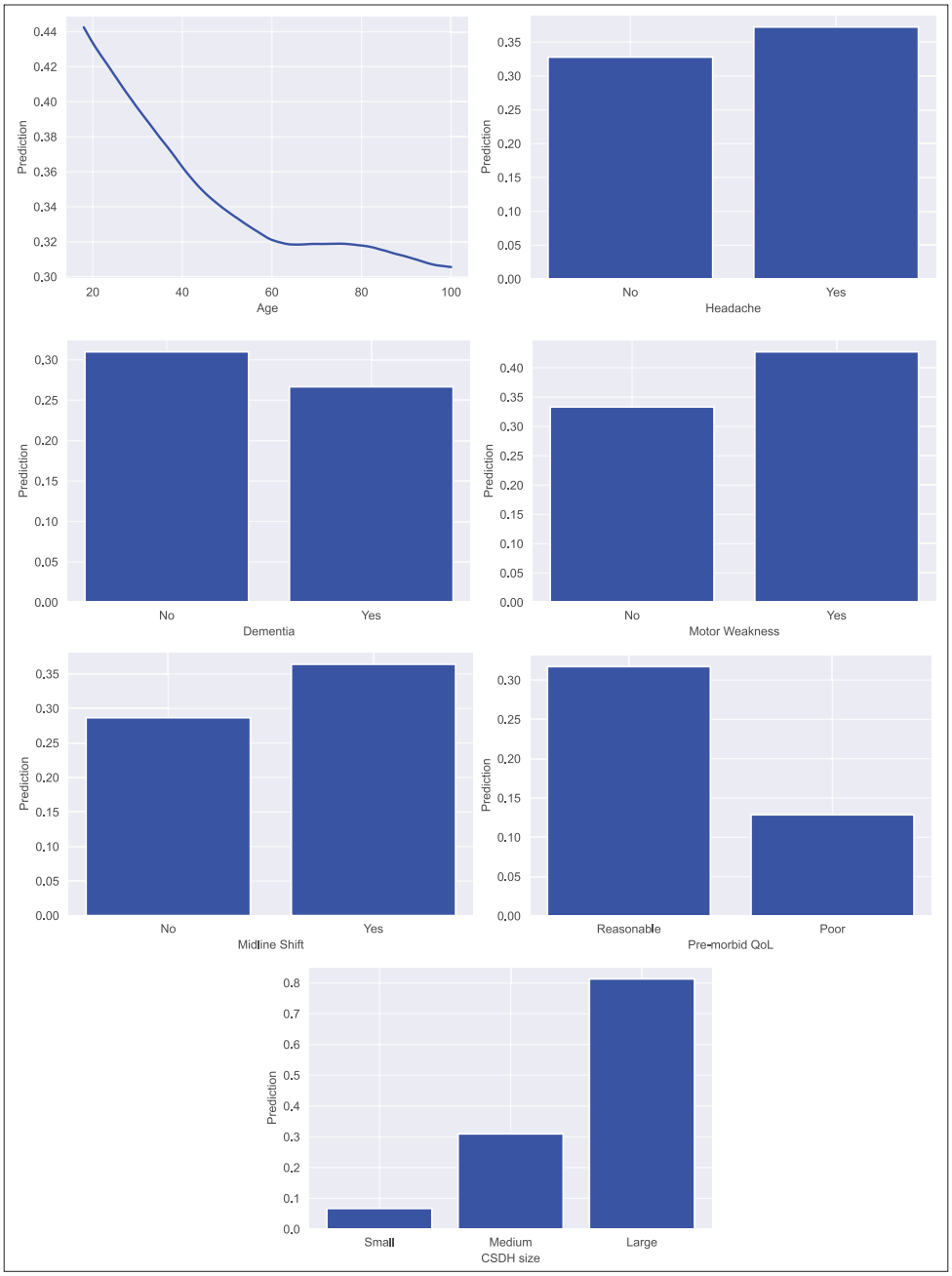
Supplementary Table 4: Cohort demographics of the external validation set with descriptive statistical analysis using *t*-tests for continuous variables (Mean [± Standard Deviation]) and Chi-square tests for categorical variables (*n* [%]).

	Rejected cohort (*n*=1208) (%)	Accepted cohort (%) (*n*=505)	Total cohort (*n*=1713) (%)	*P*-value(%)
Age	78.206±14.152	73±14	76.695±14.423	0.344
Headache				
Yes	251 (20.8)	179 (35.4)	430 (25.1)	<0.001*
No	957 (79.2)	326 (64.6)	1283 (74.9)	
Dementia				
Yes	206 (17.1)	37 (7.3)	243 (14.2)	<0.001*
No	1002 (82.9)	468 (92.7)	1470 (85.8)	
Motor weakness				
Yes	270 (22.4)	218 (43.2)	488 (71.5)	<0.001*
No	938 (77.6)	287 (56.8)	1225 (29.5)	
Midline shift				
Yes	160 (13.2)	425 (84.2)	585 (34.2)	<0.001*
No	1048 (86.8)	287 (56.8)	1128 (65.8)	
Size of CSDH				
Small	785 (65)	3 (0.6)	788 (46)	<0.001*
Medium	326 (27)	95 (18.8)	421 (24.6)	
Large	97 (8)	407 (80.6)	504 (29.4)	
Pre-morbid QoL				
Poor	462 (38.2)	12 (2.4)	474 (27.7)	<0.001*
Reasonable	746 (61.8)	493 (97.6)	1239 (72.3)	

CSDH: Chronic subdural hematoma, QoL: Quality of life. **P*-value significant at the 0.05 level

External validation of ANCHOR

The average age of the London cohort was 76.695 ± 14.423 years. Detailed baseline characteristics of the external validation patient cohort are summarized in Supplementary Table 4. Comparative analysis revealed that the baseline characteristics of the external validation cohort did not differ from those of the internal cohort for any of the 7 variables, as shown in Table 3. The analysis demonstrated no statistically significant difference in the age of the local patient cohort compared to the external validation cohort ($P = 0.089$) and in the rates of headache ($P = 0.485$),



Supplementary Figure 1: Partial dependence plots for each predictor variable on the binary outcome class for the artificial neural network model. CSDH = Chronic subdural hematoma, QoL = Quality of Life.

dementia (P = 0.833), motor weakness (P = 0.171), midline shift (P = 0.745), and pre-morbid QoL (P = 0.415). The London cohort demonstrated a comparatively greater incidence of medium sized hematomas as compared to the internal cohort; however, there was no statistically significant difference (P = 0.746). Acceptance status between the internal and external cohorts also exhibited no statistically significant difference, with 30% and 29.5% acceptance rates, respectively (P = 0.747).

The ANCHOR ML model demonstrated good discriminative ability, calibration, overall performance, and accuracy on this external validation cohort of 1713 CSDH patient referrals from London. On external validation, the algorithm had an AUC of 0.896 (95% CI: 0.878–0.912) [Figure 4], suggesting good-to-excellent discrimination. In addition, the model demonstrated an accuracy of 92.294% (95% CI: 90.952–93.520), sensitivity of 82.970 (95% CI: 79.644–86.180), specificity of 96.190% (95% CI: 94.951–97.202), and PPV of 90.108% (95% CI: 87.197–

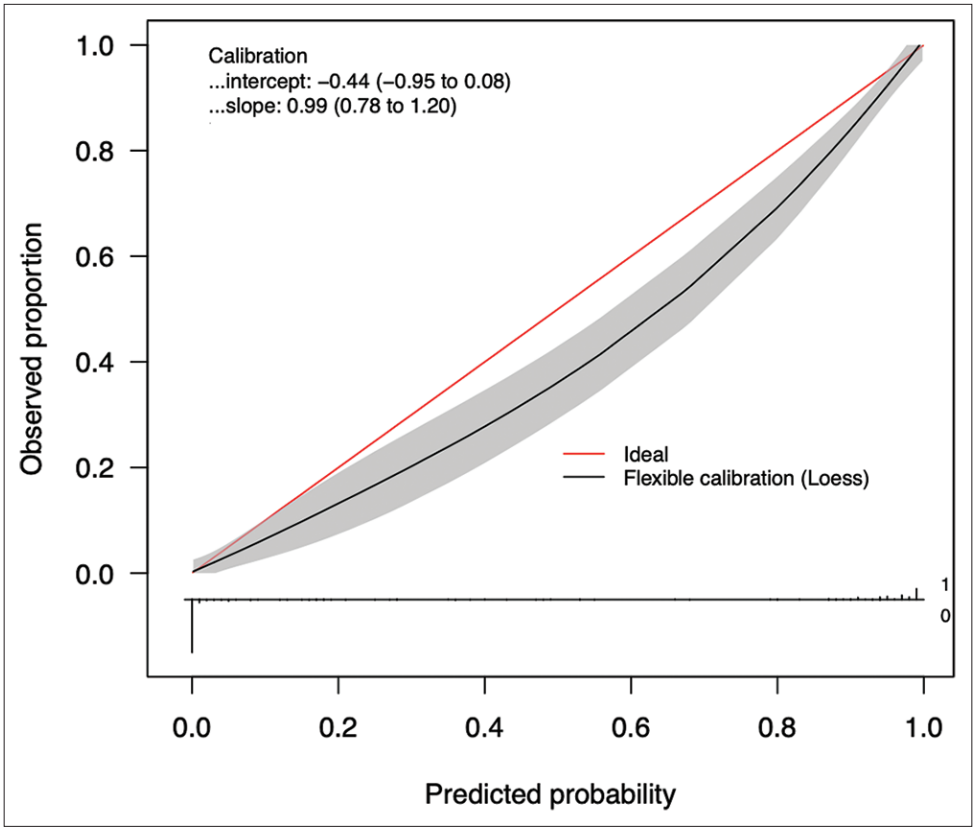


Figure 3: Calibration curve of the artificial neural network model for predicting acceptance of chronic subdural hematoma referrals on the testing set, $n = 450$.

92.688). DCA showed that ANCHOR provided greater clinical net benefit at all predicted probabilities relative to default strategies of management for all or no patients [Figure 5]. There was also excellent prediction of CSDH referral outcome from the predicted probabilities of 0.8–1.0 [Figure 6]. However, the ANCHOR algorithm underestimated the observed proportion of patients with referral acceptance, with predicted probabilities lower than 0.8. This finding is represented by an overall calibration intercept of 0.34 (0.13–0.55), an overall calibration slope of 0.92 (0.83–1.01) and a brier score loss of 0.077 (95% CI: 0.065–0.091).

Feature analysis

RFE analysis also revealed that the 7 aforementioned predictor variables were necessary to achieve the highest average accuracy. Model agnostic SHAP feature importance calculations revealed the size of the CSDH to be the single most important predictor of referral acceptance, with a mean SHAP value of 0.30. Please refer to Figure 7 for the SHAP feature importance scores for each of the remaining predictor variables. In addition, the direction of effect of these variables on the outcome classes is illustrated by PDPs in Supplementary Figure 1. In addition, the website illustrates locally interpretable model agnostic explanations for each set of inputs for an individual patient, highlighting the predictive impact of each variable on the result in real time.

DISCUSSION

To the best of our knowledge, this is the first study that has analyzed the CSHD referral acceptance process through ML algorithms. The increasing rate of CSDH incidence and referrals compounded with the lack of objective clinical reasoning guidelines, necessitates the need for a more accurate and efficient referral decision-making process, established on evidence-based predictions that are reliable, consistent and can take place in real time, making this process an ideal candidate for ML modeling. This study thus evaluated the utility and efficacy of multiple ML models in predicting the outcome of a CSDH patient referral. The best performing model was validated on an external cohort successfully demonstrating excellent accuracy in predicting CSDH referral outcome. Used as a clinical triage and decision support tool, our model could thus ensure that those most likely to need surgical intervention are assessed first.

We report that our ANN model demonstrated the best overall discriminatory ability with good calibration upon validation of unseen data points. Such conventional black box models are well

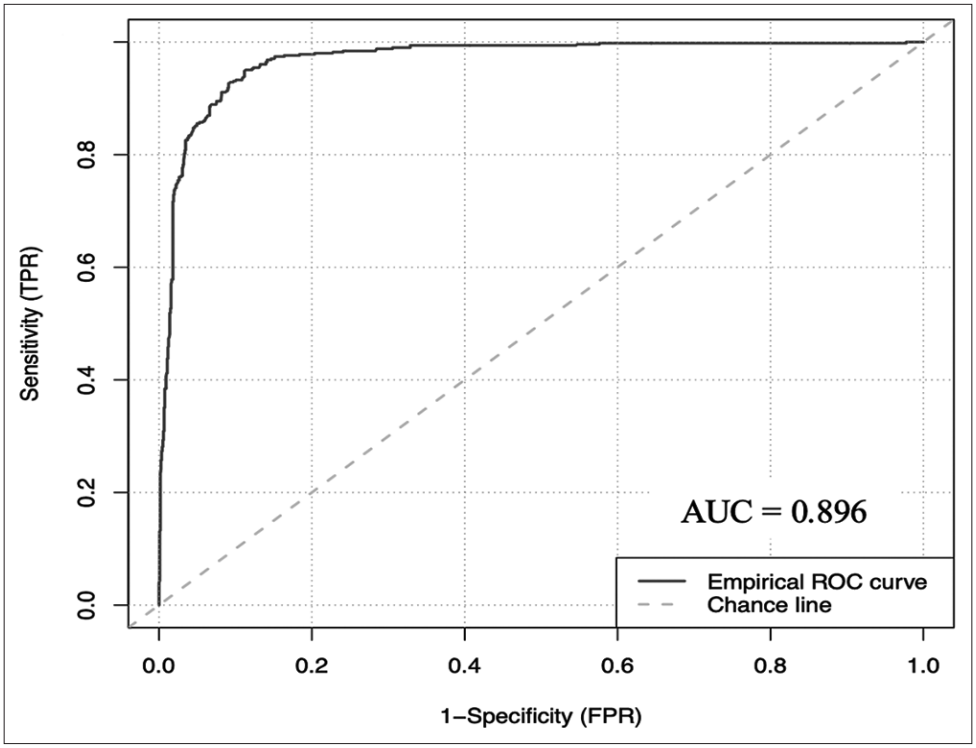


Figure 4: Receiver operating curve of the artificial neural network model for predicting acceptance of chronic subdural hematoma referrals, AUC = 0.896, on the external validation set, *n* = 1713. AUC = Area under the receiver operating curve.

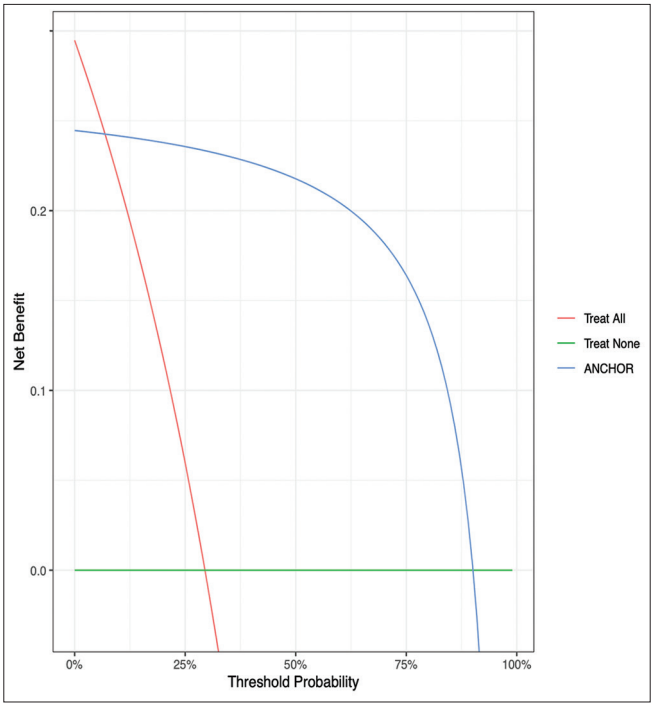


Figure 5: Decision curve analysis comparing expected clinical net benefit of the 5 different machine learning models on the external validation set. SVM = Support Vector Machine, KNN = K-nearest neighbours, DT = Decision Tree, ANN = Artificial Neural Network, ANCHOR = Artificial Neural network for Chronic subdural HematOma Referral outcome prediction.

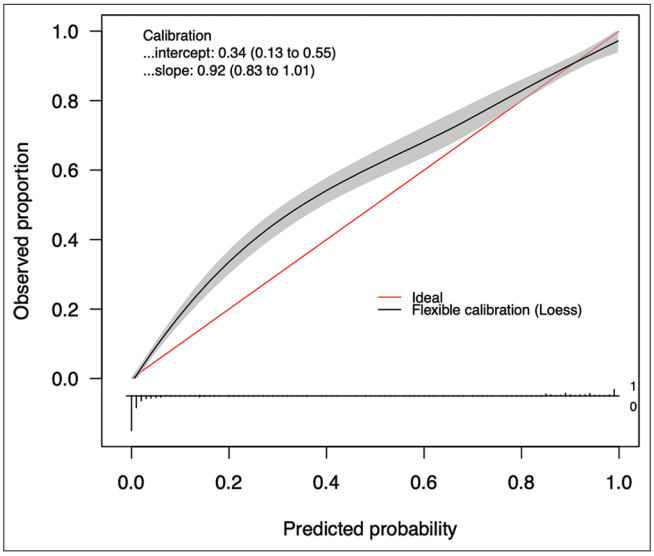


Figure 6: Calibration curve of the artificial neural network model for predicting acceptance of chronic subdural hematoma referrals on the external validation set, $n = 1713$.

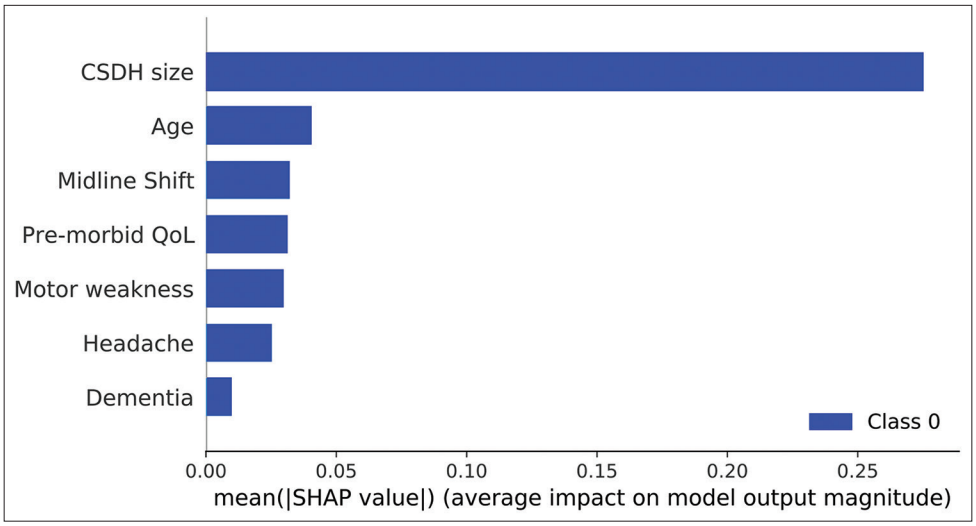


Figure 7: SHAP feature importance scores of the 7 predictor variables in the artificial neural network model. Mean |SHAP| values provide the absolute value of the average feature importance scores across all cross-validation folds. SHAP: SHapley Additive exPlanation values. CSDH: Chronic subdural hematoma, QoL: Quality of Life.

suited for analyzing complex non-linear relationships between predictor and outcome variables. However, they suffer from a lack of explain ability and interpretability. Abouzari *et al*. have previously demonstrated the superior efficacy of ANNs compared to logistic regression ML models in outcome prediction in CSDH patients; however, their study suffered from poor explain ability and had a low sample size for training and validating the model.[3] Thus, in accordance with predictive decision-making artificial intelligence guidelines,[17] our ML models were optimized to possess an accurate predictive ability and be easy to interpret. A combination of model agnostic interpretation methods and conventional statistical techniques have allowed for objective analysis of decision-making parameters and has addressed the current gaps in understanding the role of baseline patient parameters that influence patient care.

CSDH referral decision making process

We observed that the single most important predictive factor of acceptance was the size of the hematoma. Multiple studies have previously demonstrated the negative prognostic impact of subdural hematoma size on the recurrence and post-operative outcomes of patients.[24,27,42] We report that patients with larger hematomas were significantly more likely

to be accepted compared to those with smaller hematomas. Brennan *et al*. similarly report that the most common reason for rejecting a CSDH referral was that the subdural collection was small and insufficient to explain the patient's symptoms.[10] The current study sub-classified the size of the hematoma into three categories via analysis of the maximal subdural thickness and degree of midline shift reported by the neuro-radiologists. However, discrepancies might arise due to the heterogeneity of the radiological reports as no specific quantifiable volumetric measurements were used. Kellogg *et al*. have developed 3D convolutional neural networks capable of automated segmentation and analysis of CSDH computer tomography scans, laying the groundwork for future improvement and adoption of these technologies.[20]

The findings of this study also re-enforce results previously demonstrated, regarding the non-significant predictors of referral acceptance. Brennan *et al*. have shown no significant differences in the GCS scores between transferred and non-transferred patients, with majority of patients having a GCS score of 13–15.[10] Our results support these findings, with no observable difference in the median GCS score between the accepted and rejected referral groups. Additionally, the presence of headache and motor weakness were significant predictors of referral acceptance, thereby highlighting that these variables may provide a better representation of the patient's neurological status overtime compared to their GCS score at referral. Finally, in our study, around 37% of CSDH patients referred to neurosurgery while on anti-platelet and anti-coagulation medication were accepted. Although being on antithrombotic agents increases the risk of subdural recollection after surgery,[41] we found that the decision making process for the index operation was not altered by the patient being on them. These results are in line with the previous literature,[10] thereby further emphasizing the generalizability of our model to a wider population group.

Goals of ANCHOR

Across the UK's two largest tertiary neurosurgical centers that provide care to a large and diverse population of patients, we have demonstrated the capability of ANCHOR as a decision making adjunct within the health-care framework of the UK.

Focus was placed on the transparency, interpretability and explain ability of the model for both patients and physicians. Such methods allow patients to understand what is happening and why the decision to accept or reject their referral is being made. In addition, these features combined with the real time output of referral outcome allow for faster and safer transmission of information across all involved medical professionals. This tool can also be useful to the junior physicians in training, who can check whether their initial management plan is sound. At present, acceptance of CSDH referrals to neurosurgery is a decision influenced by an individual clinician's experience and is based on the impact of patient variables such as age, size of the subdural hematoma, and neurological deficit.[22,32] Thus, this open access model will provide evidence based knowledge and information to our referring non-neurosurgical colleagues regarding the nuances of the referral decision making process, thereby only allowing for specific and appropriate patient referrals.

Finally, the UK NHS is currently undergoing a digital health-care revolution, with a focus on the creation, distribution and enhancement of reliable health-care systems and research pipelines using information technology and AI.[13,18] In support, this pilot study has addressed its primary aim by creating a reliable decision-support ML tool and publicly deploying it for testing by healthcare professionals and thus is well suited to facilitate the development of more sophisticated decision tools for CSDH patients in the future. This study also shows that ML algorithms of this nature could be used for other clinical conditions with multiple variables determining binary outcomes.

Limitations

Despite these results, our study has a few limitations. First, the decision of the consultant neurosurgeon was assumed to be correct and no follow-up data were collected on the outcomes of these decisions. Second, the predictor variables included in this analysis are not exhaustive. Factors such as race, ethnicity, co-morbidities (diabetes, cardiovascular disease, etc.), the presence of previous hematomas, volumetric radiological hematoma appearance, and other neurological symptoms such as ataxia and hemisensory loss were not assessed. These variables can possibly influence the outcomes of the referrals. Third, while our model demonstrated excellent accuracy across our internal and external cohorts, the results cannot be generalized to the CSDH patient population worldwide. Countries with a different funding model of healthcare and less centralization of neurosurgical services may not find this tool helpful. Therefore, external validation in multiple international tertiary centers in differing healthcare settings is thus necessary to safely employ the model in clinical practice. Despite the differences in neurosurgical practices worldwide, there is however a global applicability of such a publicly available ML model as it can be used by non-specialist clinicians who encounter a CSDH and are debating whether they need to refer this onto neurosurgery. In addition, our work shows that if there is good data already available on the referrals, a ML algorithm can be built specifically for that country or region to aid the management of this condition. Fourth, the small incongruities in our model's calibration on the internal and external datasets can be explained by the cumulative yet non-significant differences in the distribution of patient variables. These changes may represent a "gray zone" of variable clinical presentations that require more

in-depth individualistic evaluation and thus, may not be well evaluated by our ML model. This suggests that further prospective training and enhancement are required before this tool is used in routine clinical practice.

CONCLUSION

This is the first study to evaluate the use of ML algorithms in deciding the outcome of a referral for CSDH and demonstrates that implementation of accurate and reliable ML algorithms as decision support tools is feasible and can potentially be used in conjunction to current clinical practice. In addition, the study highlights which variables influence clinical decision making the most. These findings can help facilitate the CSDH clinical referral decision making process and have the potential to significantly enhance patient care and physician education.

Acknowledgments

The authors wish to thank: (1) Callum Tetlow, Data Scientist, at Northern Care Alliance, for statistical and computational input and analysis and (2) the department of neurosurgery at Salford Royal Hospital for providing the data.

Declaration of patient consent

Patients' consent not required as patients' identities were not disclosed or compromised.

Financial support and sponsorship

Nil.

Conflicts of interest

There are no conflicts of interest.

REFERENCES

1. Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, *et al*. TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems. arXiv 2016;2:1-19.
2. Abadi M, Barham P, Chen J, Chen Z, Davis A, Dean J, *et al*. TensorFlow: A System for Large-scale Machine Learning. In: Open Access to the Proceedings of the USENIX Symposium on Operating. Systems Design and Implementation. Berkeley: USENIX Association; 2016. p. 265-83.
3. Abouzari M, Rashidi A, Zandi-Toghani M, Behzadi M, Asadollahi M. Chronic subdural hematoma outcome prediction using logistic regression and an artificial neural network. Neurosurg Rev 2009;32:479-84.
4. Adhiyaman V, Chattopadhyay I, Irshad F, Curran D, Abraham S. Increasing incidence of chronic subdural haematoma in the elderly. QJM 2017;110:375-8.
5. Asghar M, Adhiyaman V, Greenway MW, Bhowmick BK, Bates A. Chronic subdural haematoma in the elderly--a North Wales experience. J R Soc Med 2002;95:290-2.
6. Aspegren OP, Åstrand R, Lundgren MI, Romner B. Anticoagulation therapy a risk factor for the development of chronic subdural hematoma. Clin Neurol Neurosurg 2013;115:981-4.
7. Berrar D. Cross-validation. In: Encyclopedia of Bioinformatics and Computational Biology. Netherlands: Elsevier; 2019. p. 542-5.
8. Biswas S, MacArthur J, Sarkar V, Thompson H, Saleemi M, George KJ. Development and validation of the chronic subdural hematoma referral outcome prediction using statistics (CHORUS) score: A retrospective study at a national tertiary centre. World Neurosurg 2022;In Press:S1878875022016497.
9. Bradley AP. The use of the area under the ROC curve in the evaluation of machine learning algorithms. Pattern Recognit 1997;30:1145-59.
10. Brennan PM, Kolias AG, Joannides AJ, Shapey J, Marcus HJ, Gregson BA, *et al*. The management and outcome for patients with chronic subdural hematoma: A prospective, multicenter, observational cohort study in the United Kingdom. J Neurosurg 2017;127:732-9.
11. Buitinck L, Louppe G, Blondel M, Pedregosa F, Mueller A, Grisel O, *et al*. API Design for Machine Learning Software: Experiences from the Scikit-learn Project. In: European Conference on Machine Learning and Principles and Practices of Knowledge Discovery in Databases; 2013.
12. Collins GS, Reitsma JB, Altman DG, Moons K. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD statement. BMC Med 2015;13:1.
13. Cresswell K, Williams R, Sheikh A. Bridging the growing digital divide between NHS England's hospitals. J R Soc Med 2021;114:111-2.
14. Foelholm R, Waltimo O. Epidemiology of chronic subdural haematoma. Acta Neurochir (Wien) 1975;32:247-50.
15. Gneiting T, Raftery AE. Strictly proper scoring rules, prediction, and estimation. J Am Stat Assoc 2007;102:359-78.
16. Greenwell BM, Boehmke BC, McCarthy AJ. A Simple and Effective Model-Based Variable Importance Measure. ArXiv Stat ML 2018; 1:1-27.
17. de Hond AA, Leeuwenberg AM, Hooft L, Kant IM, Nijman SW, van Os HJ, *et al*. Guidelines and quality criteria for artificial intelligence-based prediction models in healthcare: A scoping review. NPJ Digit Med 2022;5:2.
18. Joshi I, Morley J. Artificial Intelligence: How to get it Right. Putting Policy into Practive for Safe Data-driven Innovation in Health and Care. London, United Kingdom: NHSX; 2019. p. 6-106.
19. Junker M, Hoch R, Dengel A. On the Evaluation of Document Analysis Components by Recall, Precision, and Accuracy. In: Proceedings of the Fifth International Conference on Document Analysis and Recognition. ICDAR '99 (Cat. No.PR00318). Bangalore, India: IEEE; 1999. p.713-6.
20. Kellogg RT, Vargas J, Barros G, Sen R, Bass D, Mason JR, *et al*. Segmentation of chronic subdural hematomas using 3D convolutional neural networks. World Neurosurg 2021;148:e58-65.
21. Kohavi R. A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. In: Proceedings

of the 14th International Joint Conference on Artificial Intelligence. Vol. 2. IJCAI'95. San Francisco, CA: Morgan Kaufmann Publishers Inc.; 1995. p. 1137-43.
22. Kolias AG, Chari A, Santarius T, Hutchinson PJ. Chronic subdural haematoma: Modern management and emerging therapies. Nat Rev Neurol 2014;10:570-8.
23. Kudo H, Kuwamura K, Izawa I, Sawa H, Tamaki N. Chronic subdural hematoma in elderly people: Present status on Awaji Island and epidemiological prospect. Neurol Med Chir (Tokyo) 1992;32:207-9.
24. Leroy HA, Aboukaïs R, Reyns N, Bourgeois P, Labreuche J, Duhamel A, *et al*. Predictors of functional outcomes and recurrence of chronic subdural hematomas. J Clin Neurosci 2015;22:1895-900.
25. Luo W, Phung D, Tran T, Gupta S, Rana S, Karmakar C, *et al*. Guidelines for developing and reporting machine learning predictive models in biomedical research: A multidisciplinary view. J Med Internet Res 2016;18:e323.
26. Marcilio WE, Eler DM. From Explanations to Feature Selection: Assessing SHAP Values as Feature Selection Mechanism. In: 2020 33rd SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI). Brazil: IEEE: Recife/Porto de Galinhas; 2020. p. 340-7.
27. Miah IP, Tank Y, Rosendaal FR, Peul WC, Dammers R, Lingsma HF, *et al*. Radiological prognostic factors of chronic subdural hematoma recurrence: A systematic review and meta-analysis. Neuroradiology 2021;63:27-40.
28. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, *et al*. Scikit-learn: Machine learning in python. J Mach Learn Res 2011;12:2825-30.
29. Rauhala M, Luoto TM, Huhtala H, Iverson GL, Niskakangas T, Öhman J, *et al*. The incidence of chronic subdural hematomas from 1990 to 2015 in a defined Finnish population. J Neurosurg 2020;132:1147-57.
30. Rust T, Kiemer N, Erasmus A. Chronic subdural haematomas and anticoagulation or anti-thrombotic therapy. J Clin Neurosci 2006;13:823-7.
31. Santarius T, Kirkpatrick PJ, Kolias AG, Hutchinson PJ. Working toward rational and evidence-based treatment of chronic subdural hematoma. Clin Neurosurg 2010;57:112-22.
32. Sharma R, Rocha E, Pasi M, Lee H, Patel A, Singhal AB. Subdural hematoma: Predictors of outcome and a score to guide surgical decision-making. J Stroke Cerebrovasc Dis 2020;29:105180.
33. Sidey-Gibbons JA, Sidey-Gibbons CJ. Machine learning in medicine: A practical introduction. BMC Med Res Methodol 2019;19:64.
34. Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: Seven steps for development and an ABCD for validation. Eur Heart J 2014;35:1925-31.
35. Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, *et al*. Assessing the performance of prediction models: A framework for traditional and novel measures. Epidemiol Camb Mass 2010;21:128-38.
36. Toi H, Kinoshita K, Hirai S, Takai H, Hara K, Matsushita N, *et al*. Present epidemiology of chronic subdural hematoma in Japan: Analysis of 63,358 cases recorded in a national administrative database. J Neurosurg 2018;128:222-8.
37. Trevethan R. Sensitivity, specificity, and predictive values: Foundations, pliabilities, and pitfalls in research and practice. Front Public Health 2017;5:307.
38. Uno M, Toi H, Hirai S. Chronic Subdural Hematoma in Elderly Patients: Is this Disease Benign? Neurol Med Chir (Tokyo) 2017;57:402-9.
39. Van Calster B, McLernon DJ, van Smeden M, Wynants L, Steyerberg EW. Calibration: The Achilles heel of predictive analytics. BMC Med 2019;17:230.
40. Vickers AJ, van Calster B, Steyerberg EW. A simple, step-by-step guide to interpreting decision curve analysis. Diagn Progn Res 2019;3:18.
41. Wang H, Zhang M, Zheng H, Xia X, Luo K, Guo F, *et al*. The effects of antithrombotic drugs on the recurrence and mortality in patients with chronic subdural hematoma: A meta-analysis. Medicine (Baltimore) 2019;98:e13972.
42. Yamamoto H, Hirashima Y, Hamada H, Hayashi N, Origasa H, Endo S. Independent predictors of recurrence of chronic subdural hematoma: Results of multivariate analysis performed using a logistic regression model. J Neurosurg 2003;98:1217-21.
43. Yin M, Vaughan JW, Wallach H. Understanding the Effect of Accuracy on Trust in Machine Learning Models. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. Glasgow Scotland UK: ACM; 2019. p. 1-12.

How to cite this article: Biswas S, MacArthur JI, Pandit A, McMenemy L, Sarkar V, Thompson H, *et al*. Predicting neurosurgical referral outcomes in patients with chronic subdural hematomas using machine learning algorithms – A multi-center feasibility study. Surg Neurol Int 2023;14:22.