RESEARCH ARTICLE

A novel toolkit based on Artificial Intelligence and Automated Machine Learning to aid outcome prediction and decision-making in the selection of patients undergoing Mechanical Thrombectomy in Acute Ischaemic Stroke

Sanjeev Nayak MRCP, FRCR[1*], Vijay Jeganath FRCA, FFICM[2] and Qian Zhao PhD[3]

[1]Consultant Interventional Neuroradiologist, University Hospitals of North Midlands NHS Trust, Stoke-on-Trent, UK
[2]Consultant Anaesthetist, University Hospitals of North Midlands NHS Trust, Stoke-on-Trent, UK (Presently, Associate Director of Medical Education, Hamad Medical Corporation, Qatar)
[3]Data Scientist, DataRobot Inc, London, UK

* sanjeev.nayak@uhnm.nhs.uk

ABSTRACT

Background: Stroke is a major cause of death and disability. Acute ischaemic strokes in selected patients are effectively treated by mechanical thrombectomy (MT). The success of MT treatment could be best estimated using highly accurate outcome prediction models which are yet to be established. The aim of this study is to develop an artificial intelligence (AI) based automated machine learning (AutoML) toolkit to aid decision-making for mechanical thrombectomy based on readily available patient variables that could predict functional outcome following the treatment.

Methods: Datasets of 1097 patients from Systematic Evaluation of Patients Treated With Stroke Devices for Acute Ischemic Stroke (STRATIS) Registry and Solitaire™ with the Intention for Thrombectomy as Primary Endovascular Treatment for Acute Ischemic Stroke (SWIFT PRIME) Trial were retrospectively evaluated. Linear and non-linear models were built using an automated machine learning platform, DataRobot. We developed two stage models for predicting the outcome of the patient:
Model 1 predicted survival, defined as an mRS score of 0-5 (alive) or 6 (dead).
Model 2 predicted good/bad survivor, defined as an mRS score of 0-2 (good) or 3-5 (poor).

Results: The primary outcome was the modified Rankin Scale (mRS) score at 90 days after stroke. Predictions were measured by area under curve (AUC). Prediction for survival was 83% accurate (AUC 0.7780). Prediction of good/poor survival was 61% accurate (AUC 0.7061). A two-stage machine learning model has an improved 82% overall accuracy of prediction.

Conclusion: The proposed AI-based AutoML toolkit evaluates various baseline clinical and radiological characteristics and predicts significant variations in treatment benefit between patients. With its improved prediction accuracy, the toolkit is clinically useful as it helps in distinguishing between individual patients who may experience benefit from mechanical thrombectomy treatment for acute ischaemic stroke from those who may not.

Keywords: Acute ischaemic stroke; Mechanical thrombectomy (MT); Large vessel occlusion (LVO); Artificial Intelligence (AI); Automated Machine Learning (AutoML); Prediction scoring system.

INTRODUCTION

Stroke is a major cause of mortality and morbidity. It is the second commonest cause of death globally, and the third commonest cause of death in the UK.[1,2] In Western countries, 80% of strokes are ischaemic. The ischaemic strokes caused by a proximal occlusion in the intracranial cerebral arteries result in poor outcome[3,4].

Mechanical Thrombectomy (MT) is a highly successful procedure that improves functional outcome in patients with acute ischaemic stroke caused by a proximal occlusion, with a number needed to treat of 5 (odds ratio 2.35, 95% confidence interval 1.85 to 2.98)[5].

The concept of outcome predictors in stroke endovascular field mostly evolved from uncontrolled single-arm studies which later became an integral part of treatment selection criteria and incorporated into clinical guidelines leading to the adoption of over simplistic binary behaviours around isolated variables such as baseline stroke severity (eg, National Institutes of Health Stroke Scale [NIHSS] <6 versus ≥6), infarct burden (Alberta Stroke Program Early CT Score [ASPECTS] <6 versus ≥6), and time (≤6 versus >6 hours).[6]

The cardiologists had the benefit of comparing how treated versus untreated patients behaved across distinct variables and as such they could develop selection paradigms centred on treatment effect modifiers rather than the more restrictive predictors of outcomes.[7]

Hence it might be time to reflect on the domineering effect of medical nihilism we currently live in and act prudently to avoid the untoward consequences surrounding the over-strict interpretation of randomized clinical trials (RCTs).[7] We must recognize that we may actively harm patients by depriving them from an effective treatment just as much as we can by exposing them to a hazardous one.[7] We have developed a tool to aid the provision of MT to a greater proportion of patients who may benefit and to caution against MT in those who may not.

Despite the clinical evidence supporting MT, optimised selection methods in patients with large vessel occlusion stroke are yet to be established.[8] Although current guidelines provide some selection criteria, accurate predictive models would be useful to further refine these criteria and to aid the physician in providing the best estimate of treatment success for each patient based on dynamic interaction of variables rather than restrictive predictors of outcome which can be achieved through machine learning methods.

There have been attempts to combine several prognostic factors and predict the clinical outcome of large patients before treatment with MT. These include the Pittsburgh Response to Endovascular Therapy (PRE) score, the Stroke Prognostication Using Age and National Institutes of Health Stroke Scale (SPAN) index, the Totaled Health Risks in Vascular Events (THRIVE) score, the Houston Intra-Arterial Therapy (HIAT) score, and the HIAT2 score.[9]

These prediction models select some prognostic variables, scale or stratify them, and sum them into a univariate score, which can be called a simplified logistic regression model. This approach assumes of a linear relationship between variables and the log odds of outcomes, and it is weak to collinearity between the variables[9]. By contrast, AI through machine learning methods can produce more accurate predictive models than traditional statistical regression methods because they are more flexible and rely less on statistical assumptions than traditional regression methods[10].

We developed and validated AI based machine learning toolkit to provide individualised predictions of the effect of MT treatment based on multiple characteristics. Such a tool may be helpful to support clinical judgment when making complicated decisions on MT treatment.

METHODS
Patient Population

Datasets of 1097 patients from Systematic Evaluation of Patients Treated with Stroke Devices for Acute Ischemic Stroke (STRATIS) Registry and Solitaire™ with the Intention for Thrombectomy as Primary Endovascular Treatment for Acute Ischemic Stroke (SWIFT PRIME) Trial were retrospectively evaluated. All patients with acute ischemic stroke caused by LVO of the anterior circulation received emergent endovascular recanalization therapy. The exclusion criteria were patients with LVO of the posterior circulation.

Patient characteristics

The patient characteristics of the study population (n = 1,097) are shown in Table 1. The mean age was 68 ± 14 years. The rates of atrial fibrillation, diabetes, and hypertension were 38%, 24%, and 71%, respectively.

Feature Name	Var Type	Unique	Missing	Mean	Std Dev	Median	Min	Max	Target Leakage
ASPECTS	Categorical	12	0	N/A	N/A	N/A	N/A	N/A	Low
Diabetes mellitus	Categorical	2	0	N/A	N/A	N/A	N/A	N/A	Low
Occlusion location	Categorical	6	9	N/A	N/A	N/A	N/A	N/A	Low
mRS at day 90 0_2 good0	Numeric	2	0	0.45	0.5	0.0	0.0	1.0	N/A
NIHSS at baseline	Numeric	23	0	17.28	5.35	17.0	8.0	30.0	Low
Age	Numeric	73	1	68.013	14.043	69.0	19.0	100.0	Low
Collateral grade	Categorical	6	0	N/A	N/A	N/A	N/A	N/A	Low
Hypertension	Categorical	2	0	N/A	N/A	N/A	N/A	N/A	Low

Table 1: Outputs of exploratory data analysis from DataRobot

Model development

Patient characteristics obtained before treatment that are expected to predict outcome or to interact with treatment, based on expert opinion or recent literature, were specified in advance in our statistical analysis plan.[5,10]

Data on 10 variables were collected before the decision was made to perform MT. They included demographic data, such as the patient's age, sex, and comorbidity (hypertension, diabetes mellitus, previous stroke and atrial fibrillation). Clinical variables, such as the neurological severity measured by NIHSS, the site of occlusion detected on pre-treatment CT angiogram, CT ASPECTS and collateral grading was also included.

The linear and non-linear models have been built using an automated ML platform, DataRobot[11]. More than 1,000 procedure sets of data processing, feature engineering, and ML algorithm, including Support Vector Machine, Elastic Net Classifier, Regularized Logistic Regression, Stochastic Gradient Descent Classifier, Neural Network Classifier, etc., are developed from its repository. The software automatically chooses and executes suitable procedure sets when investigating the patterns in data. All the developed models were verified by cross-validation and sorted by the selected evaluation metric, e.g., the area under the curve (AUC).

Two stage model

We developed a two-stage model for the use of AI application **for** predicting the outcome of the patient:
Model 1 predicted *survival*, defined as an mRS score of 0-5 (alive) or 6 (dead).
Model 2 predicted *good or bad survival*, defined as an mRS score of 0-2 (good) or 3-5 (poor).

Data pre-processing

From a large number of data pre-processing approaches, the following approaches were automatically selected in the final models: imputing missing values, one- hot encoding for categorical values, standardization for numerical values, and creating new parameters by unsupervised learning of original parameters. Missing numerical values were imputed based on the medians of values in its parameters, and missing categorical values were treated as their own categorical level and given their own parameters. Categorical values were converted to many binary parameters by one-hot encoding if needed. For some models, numerical values were standardized in each parameter by subtracting the mean and dividing by the standard deviation. Moreover, some new parameters were created internally by

summarizing original parameters with an unsupervised learning method.

Model validation and External Validation

All developed models were validated by cross-validation and holdout, using the AUC of the receiver-operating characteristic (ROC) curve as the evaluation metric. Before developing models, 20% of the dataset was randomly selected as the holdout, which was never used in training or validation. This data set was used external validation of our model. The remaining data were randomly divided into five mutually exclusive folds of data, four of which were used together for training, with the final fold used for validation[12].

Models were trained five times per algorithm, with each fold used once for validation. Cross-validation scores were calculated by taking the mean of AUC of the five possible validation folds[13]. Random selection was performed in cross-validation and holdout by stratified sampling, which holds the ratio of positive and negative cases. Finally, models were validated on the holdout to demonstrate the generalization performance to new data.

Figure 1 summarizes the CV process used by DataRobot, where the blue denotes 80.0364% of the data available for training, which is then divided into 5-folds for cross-validation and red denotes the holdout sample.
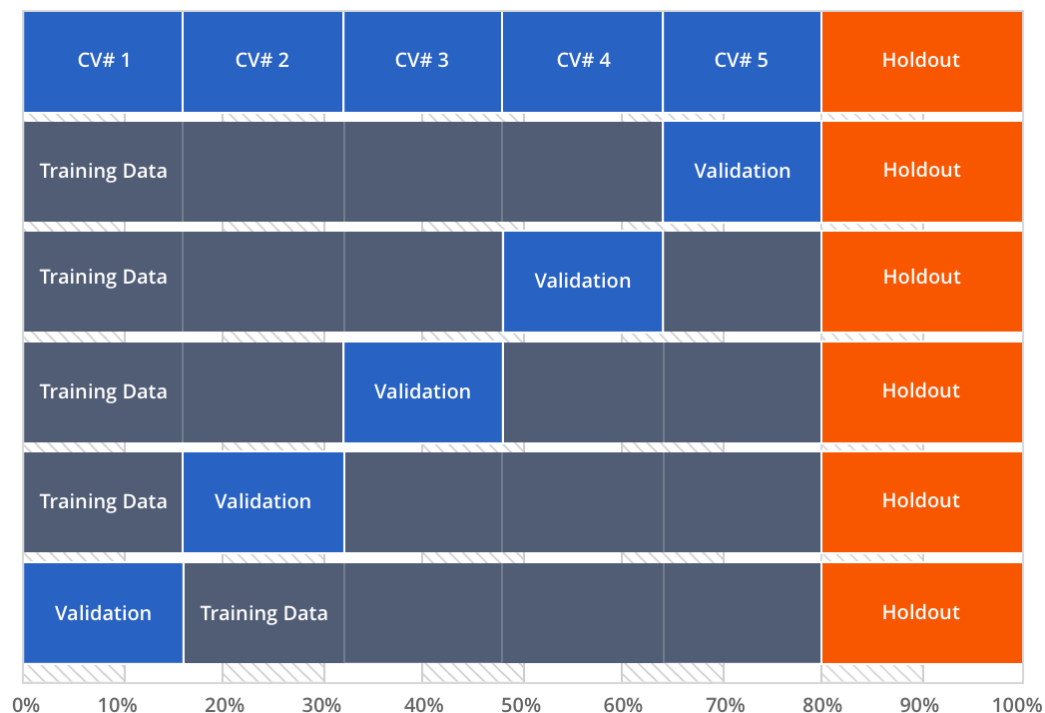

Figure 1: Cross Validation with DataRobot

DataRobot calculates the Cross Validation scores for each of the training data partitions or folds. The project metric used to calculate the score is LogLoss.

Permutation Importance

The relative importance of a parameter in the models was assessed using the permutation importance (PI), as described by Breiman[14]. This method is widely used in ML as it can be applied to both linear and non-linear models. To calculate the PI of a parameter in a model, its values in the validation data were randomly shuffled (reordered), keeping other parameters the same as

before. If it has considerable importance on the outcome, the resulting performance score in the evaluation metric should decline significantly. We calculated the PI of all parameters and divided by the maximum ratio of the resulting performance scores on the original scores to normalize and compare among different models. The calculation was conducted several times to ensure stability in random shuffling.

Partial dependence

To understand how the changes in values of a parameter affect the outcome, we constructed partial dependence plots as described by

Friedman[15]. To construct the partial dependence plot of a parameter in a model, we calculated predictions from the model after having replaced all the values for the parameter with a constant value and computing the mean of those predictions. We repeated calculations for many values to observe how the model reacts to changes in the parameter of interest.

RESULTS
Prediction of survival

Among the ML models, a model with Average Blend had the largest AUC for survival (0.7780 from all cross-validation). This model takes the predictions from several input models and averages them together into a meta-model. In this model, the top five parameters (Figure 2a) determined by PI were Age (100, reference), NIHSS at baseline (54), ASPECTS (52), Collateral grade (35) and Diabetes mellitus (10).
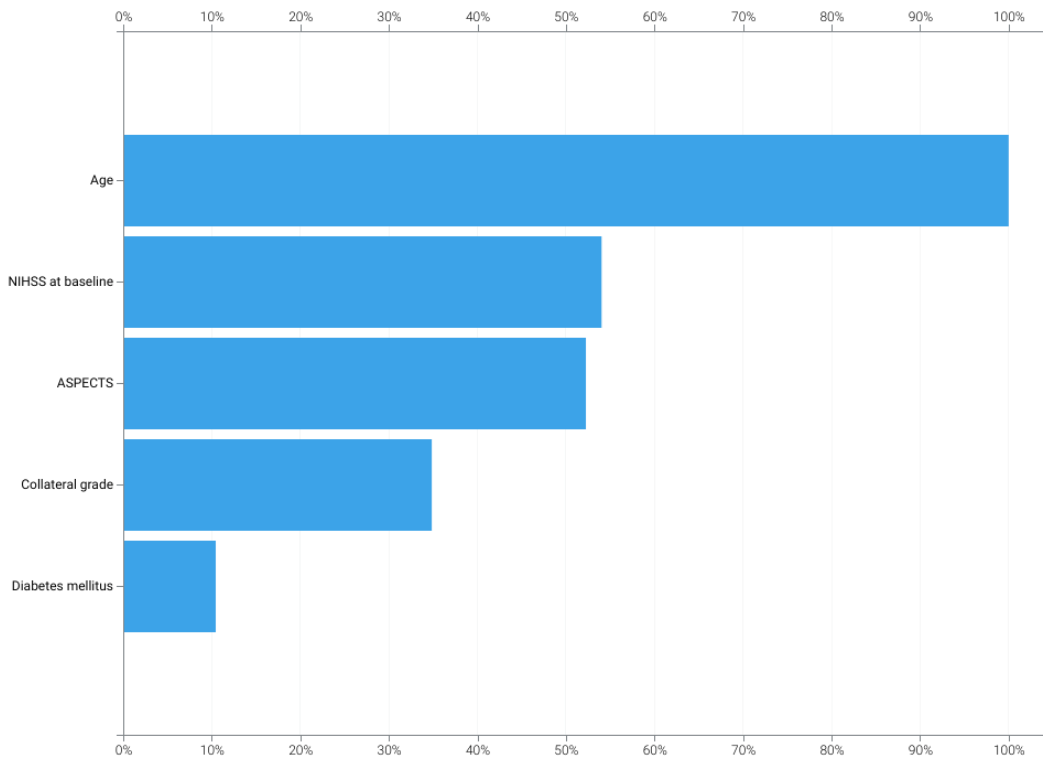

Figure 2a: Feature Impact shows, at a high level, which features are driving model decisions the most for the survival prediction model

Prediction of good/bad survival

Among the ML models, a model with Average Blend had the largest AUC for good/bad survival (0.7061 from cross-validation). This model takes the predictions from several input models and averages them together into a meta-model. In this model, the top five parameters (figure 2b) determined by PI were Collateral grade (100, reference), NIHSS at baseline (41), ASPECTS (17), Occlusion location (16) and Age (12)
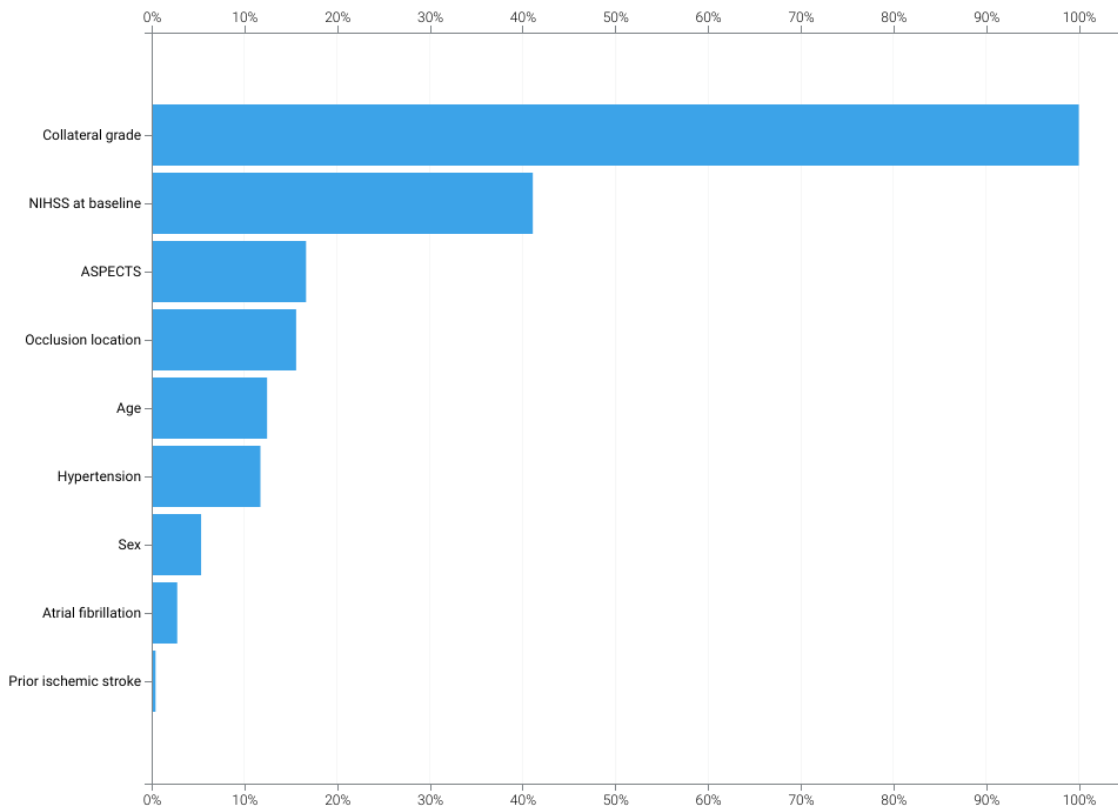
Figure 2b: Feature Impact shows, at a high level, which features are driving model decisions the most for the good/bad survivor prediction model

Two stage model prediction

With the two-stage model to predict good/bad survival, the machine learning approach has 80% accuracy for probability prediction.

For stage 1(Figure 3a), the application shows a probability score of patients having a high/low probability to survive based on variable input shown in the bottom left. In addition, the model also shows the weightage importance of individual variables (bottom right) in arriving at this conclusion.
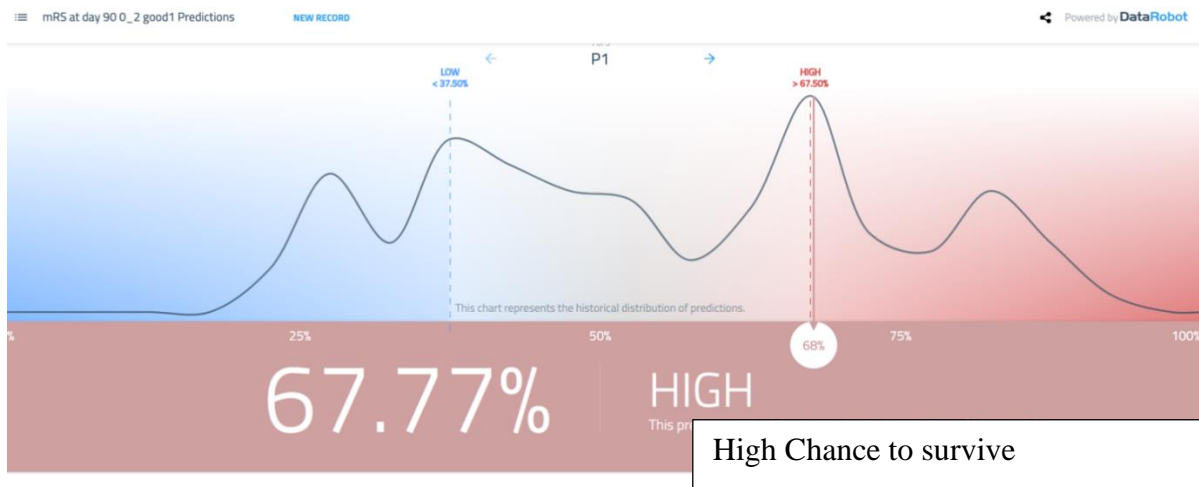
Figure 3a DataRobot AI Application makes predictions for the high/low probability to survive at patient level using any supported DataRobot model

For stage 2 (Figure 3b), the application shows a probability score of patients having a high/low probability to be a good survivor based on variable input shown in the bottom left. In addition, the model also shows the weightage importance of individual variables (bottom right) in arriving at this conclusion.
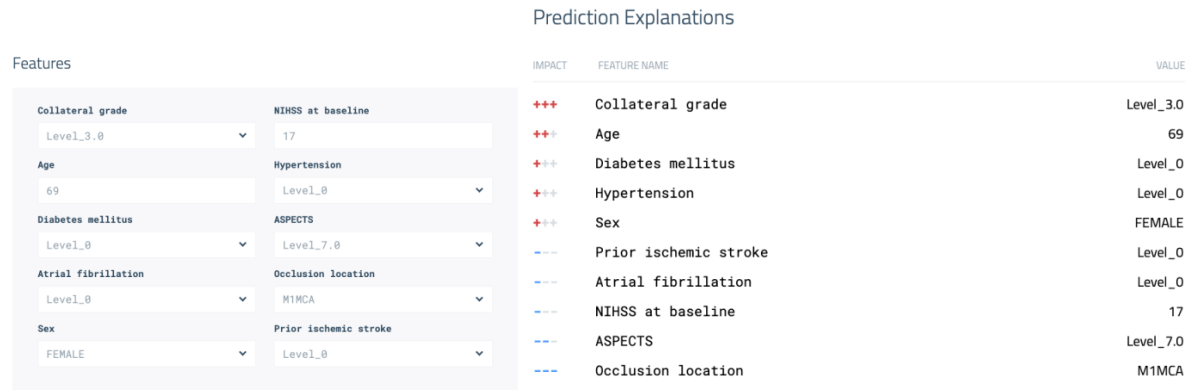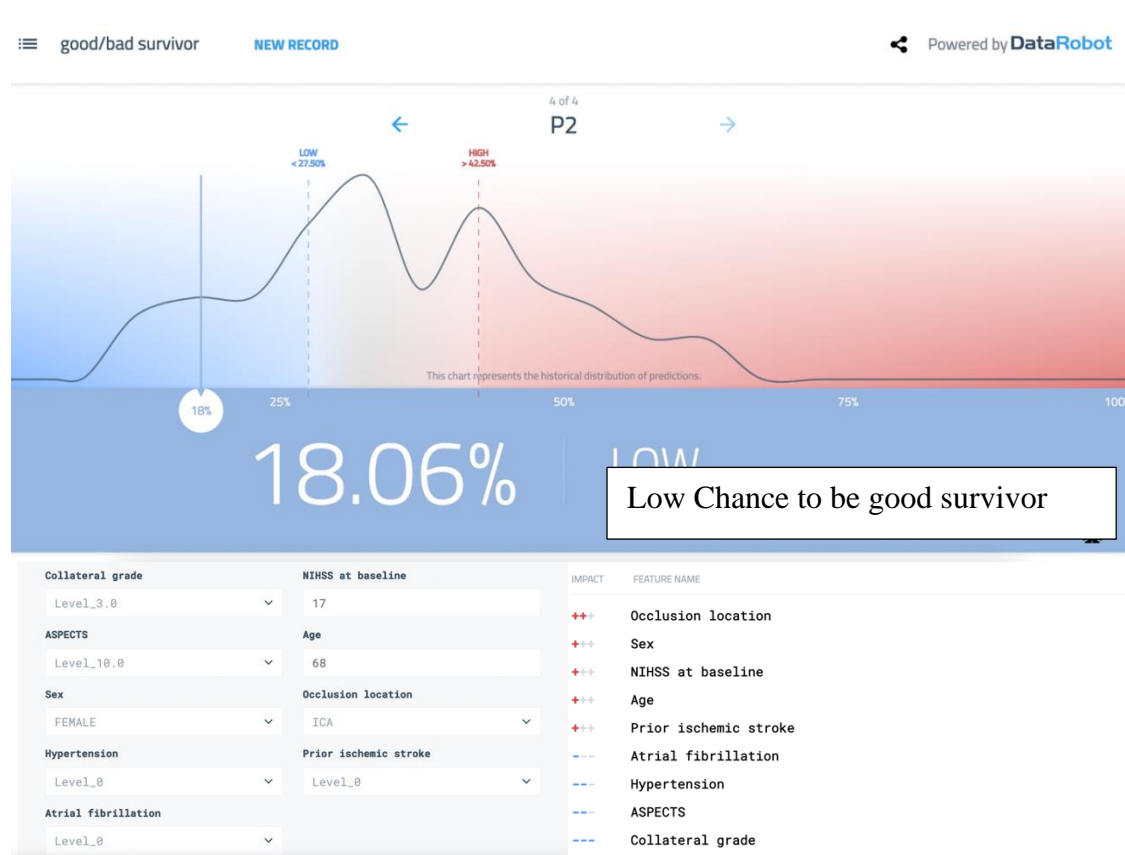

Figure 3b DataRobot AI Application makes predictions for the probability of good/bad survivor at patient level using any supported DataRobot model

DISCUSSION

Model Selection

During the model development process for survival model, we had considered the following alternative models (Table 2). The final model was selected based on model performance as well as on analysis of model diagnostics and expert business judgment.

Model Name	Validation Score	Cross Validation Score	Holdout Score	Sample Percentage
ExtraTrees Classifier (Gini)	0.3884	N/A	0.3894	63.9928
Regularized Logistic Regression (L2)	0.3655	N/A	0.3951	63.9928
Vowpal Wabbit Classifier	0.3896	N/A	0.3973	63.9928
Light Gradient Boosting on ElasticNet Predictions	0.3633	0.361	0.3974	63.9928
ENET Blender	0.3609	0.3592	0.3977	63.9928
Elastic-Net Classifier (L2 / Binomial Deviance)	0.3606	0.3594	0.3977	63.9928
Nystroem Kernel SVM Classifier	0.3612	0.3619	0.3981	63.9928
Elastic-Net Classifier (mixing alpha=0.5 / Binomial Deviance)	0.3617	0.3604	0.3982	63.9928
Elastic-Net Classifier (mixing alpha=0.5 / Binomial Deviance) with Unsupervised Learning Features	0.3605	0.3599	0.3982	63.9928

Table 2: The model types considered during the model selection process included the above models, which are sorted by the Validation score.

During the model development process for good/bad survival model, we had considered the following alternative models (Table 3). The final model was selected based on model performance as well as an analysis of model diagnostics and expert business judgment.

| Model Name	Validation Score	Cross Validation Score	Holdout Score	Sample Percentage
RandomForest Classifier (Entropy)	0.6504	N/A	0.6074	63.9928
RandomForest Classifier (Gini)	0.6495	N/A	0.6093	63.9928
Breiman and Cutler Random Forest Classifier	0.6441	N/A	0.6107	63.9928
eXtreme Gradient Boosted Trees Classifier	0.64	N/A	0.6109	63.9928
Support Vector Classifier (Radial Kernel)	0.6282	N/A	0.6111	63.9928
Gradient Boosted Trees Classifier	0.641	N/A	0.6112	63.9928
Light Gradient Boosted Trees Classifier with Early Stopping	0.6398	N/A	0.6113	63.9928
eXtreme Gradient Boosted Trees Classifier with Unsupervised Learning Features	0.6408	N/A	0.612	63.9928
Regularized Logistic Regression (L2)	0.6266	N/A	0.6143	63.9928
Generalized Additive2 Model	0.6361	N/A	0.6158	63.9928
Nystroem Kernel SVM Classifier	0.6217	0.6115	0.6162	63.9928
Advanced AVG Blender	0.6236	0.6123	0.6194	63.9928

Table 3 The model types considered during the model selection process included the above models, which are sorted by the Validation score.

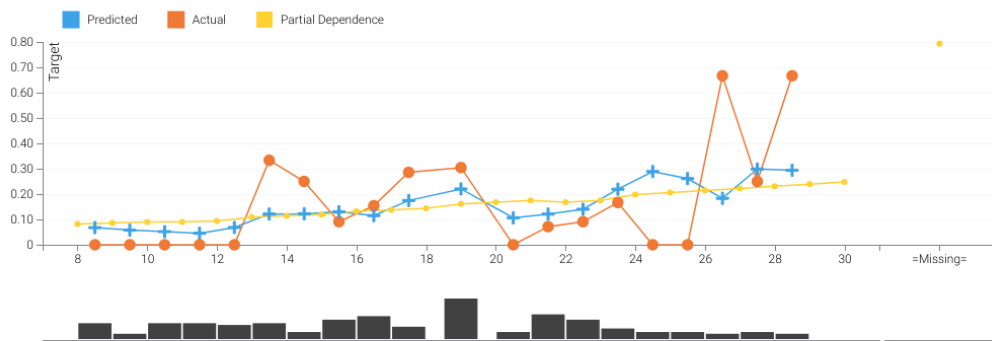**Partial dependence plots for the models**

In the case of linear regression, we can gain considerable insight into the structure and interpretation of the model by examining its coefficients. For more complex models like support vector machines, random forests, or the blenders considered here, no comparably simple parametric description is available, making the interpretation of these models more difficult. To address this difficulty for his gradient boosting machine, Friedman (2001) proposed the use of partial dependence plots (Figure 4a and 4b) . Partial dependence plots show the average partial relationship between a set of predictors and the predicted response. The partial dependence plots below capture the top features in our model, as measured by Feature Impact.

Age
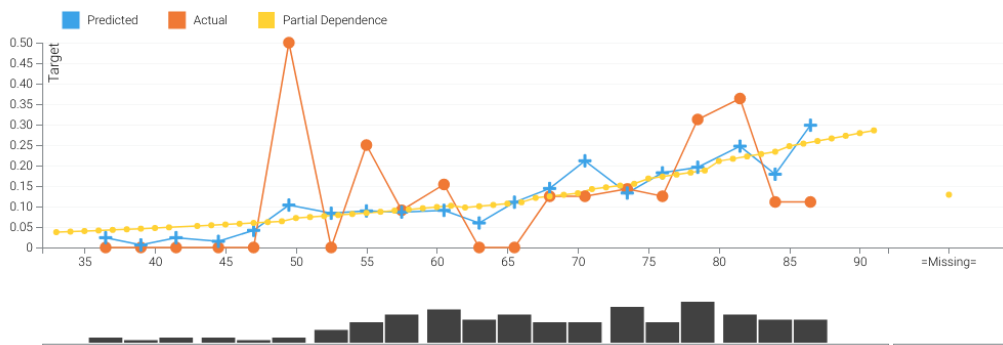
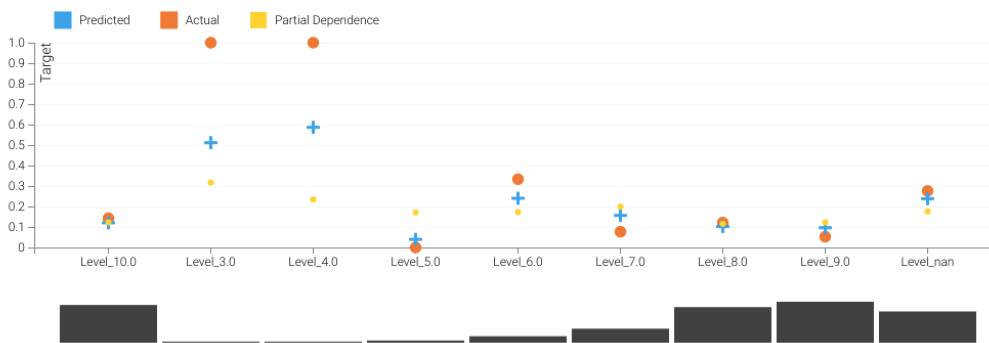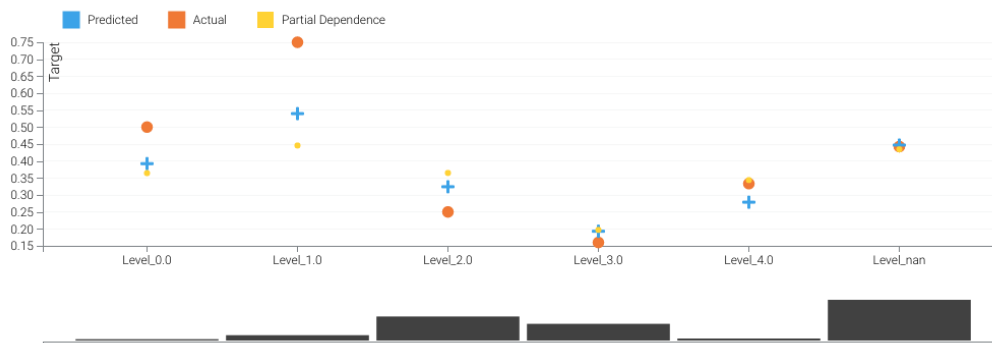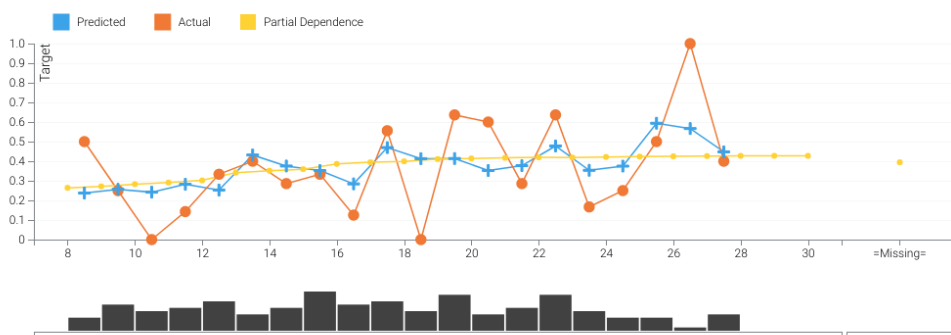## NIHSS at baseline

ASPECTS

Figure 4a: Partical dependence plots for Survival Model: Partial dependence plots capture the top features in our model, as measured by Feature Impact.

Collateral grade

NIHSS at baseline

ASPECTS
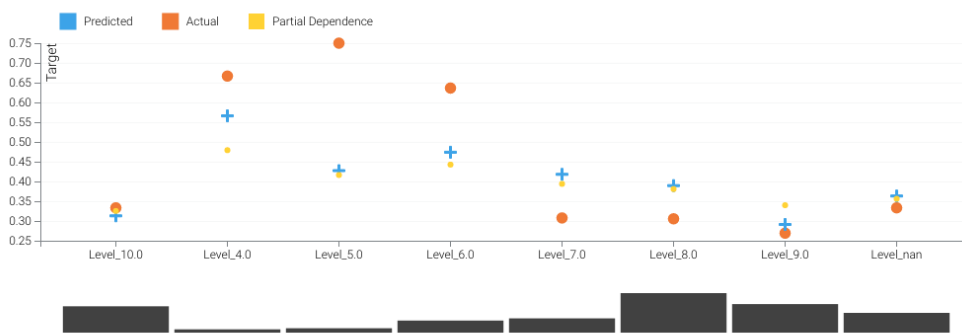
Figure 4b: Partial dependence plots for Good/Bad Survival Model. Partial dependence plots, capture the top features in our model, as measured by Feature Impact.

The orange circles depict, for the selected feature, the average target value for the aggregated feature values. The blue crosses depict, for the selected feature, the average prediction for a specific value. From the graph you can see that DataRobot also averages the predicted feature

values. Comparing the actual and predicted points can identify segments where model predictions differ from observed data. This typically occurs when the segment size is small. In those cases, for example, some models may predict closer to the overall average.

The yellow partial dependence data points depict the marginal effect of a feature on the target variable after accounting for the average effects of all other predictive features. It indicates how, holding all other variables constant, the value of this feature affects your prediction. DataRobot holds constant the values of all columns in the sample except the feature of interest. The value of the feature of interest is then reassigned to each possible value, calculating the average predictions for the sample at each setting. These values help determine how the value of each feature affects the target. The shape of the yellow data points describes the model's view of the marginal relationship between the selected feature and the target.

Two-stage model

We built a two-stage machine learning approach, with a layer of feature definitions interposed between the output of the first learned system and the input of the second. In the first stage (Table 4), the survival model is used to predict patient's probability of death undergoing MT, the output from the prediction can be survival (with prediction score 1) or non-survival (with prediction score 2), for those patients with survival as the outcome, the second-stage is used to predict good survivor (mRS 0-2 with prediction score 1A) or bad survivor (mRS 3-5 with prediction score 1B). With the two-stage model, the machine learning approach has 82% accuracy for probability prediction.

AI OUTCOME PREDICTION SCORING SYSTEM

STAGE 1

Prediction Scoring Scale	Our Scoring Category	mRS Scoring Scale	mRS Scoring Category
1	Survivor	0-5	Asymptomatic-Disabled
2	Non-survivor	6	Death

STAGE 2

Prediction Scoring Scale	Our Scoring Category	mRS Scoring Scale	mRS Scoring Category
1A	Good survivor	0-2	Asymptomatic-Mild disability
1B	Bad suvivor	3-5	Moderate-Severe disability

Table 4 The two-stage model to predict death or survival in stage 1 and good or bad survival in stage 2. The model uses a prediction scoring scale and corresponding mRS scoring scale

We further built an application that enables the physicians to start making predictions using a deployed DataRobot AI model. This application enables predictions to be made at one-at-a-time by providing the required inputs or make a large batch of predictions by importing a file. The prediction result can then be compared to the historical data from a training dataset to judge if the prediction score is high, low, or typical. This also provides the prediction explanations and adjusts input values to see how they affect the score.

Model application

Our model was developed using STRATIS Registry and SWIFT PRIME Trial database, consisting of readily available selection criteria. We evaluated a total of 1097 patient which is the largest dataset used to date to construct an intra-arterial predictive model. STRATIS had 55 study locations and SWIFT PRIME trial involved participants from United States, Austria, Denmark, France, Germany, Spain, and Switzerland. Therefore, our model covers the data characteristics from numerous stroke centres covering a large population base from a range of countries and the results of our prediction model is likely applicable

universally across all centres in Europe and the western hemisphere.

Machine learning methods can produce more accurate predictive models than traditional statistical regression methods because they are more flexible and rely less on statistical assumptions than traditional regression methods. For instance, ordinary least squares regression requires that the Gauss Markov assumptions are supported, which ensures that the model is unbiased and efficient[17].

Traditional statistical regression techniques rely on formal hypothesis testing for variable significance and feature selection (e.g., t-test, p-value, standard error). These hypothesis tests tend to have distributional and independence assumptions that may not be supported by the data. Machine learning methods, on the other hand, offer more flexibility in defining the model structure, which typically results in better model performance. Because machine learning includes methods that do not rely on formal hypothesis testing to demonstrate model validity, and because heuristic-style feature selection methods (e.g., stepwise selection) are not used in most machine learning approaches, no such distributional assumptions are required. In this case, the only assumption being made is that the model training data is representative of the future scoring data. Of course, these assumptions must be closely monitored and tracked by the model's ongoing performance monitoring process[15,16,17].

Currently, some centres withhold intra-arterial treatment in specific subgroups of patients based on imaging criteria (eg, low ASPECTS, no collaterals, or M2 occlusion)[5]. In addition, some centres also have a physician selection bias where MT is not offered based patient age (>80 years), co-morbidities such as previous stroke, anaesthetic concerns on providing a general anaesthetic or other perceived notions based on single variables where the clinician perceives that the patient may not obtain a favourable outcome.

Our study model aims to remove such perceived conceptions and shows that treatment should not be withheld based on one particular variable or a single characteristic. Some patients belonging to one of the subgroups that are considered as having no benefit of intra-arterial treatment, such as low ASPECTS, age > 80years or with co-morbidities may still benefit from intra-arterial treatment particularly if other characteristics are favourable. This emphasises the need for combining multiple clinical and radiological baseline characteristics instead of withholding treatment based on one characteristic.

Likewise, when intra-arterial treatment is not considered due to the negative effects of a general anaesthetic in a patient with severe co-morbidities, and if the prediction model shows a favourable outcome, then the procedure may be performed under local anaesthetic.

Furthermore, our model predicts no benefit of intra-arterial treatment for some patients, when more than one characteristic negatively affects the effect of intra-arterial treatment. Thus, the model may help to identify patients without expected benefit of intra-arterial treatment and topple the balance in favour of no treatment especially relevant when patients are transferred from another centre.

There are some drawbacks to the application of machine learning model in clinical practice. As the calculations cannot be performed by humans, development of software is required. Currently novel software and applications have been used in the acute ischemic stroke settings and the most well-known software is RAPID (iSchemaView, San Francisco, CA), which automatically calculates the ischemic core volume and the penumbra volume using raw magnetic resonance imaging or CT angiography data18, [18,19]. There are also applications for the image transfer, information/message sharing (chat) system. These software applications are widely used in clinical practice and in making the decision whether performing MT is supported, enhanced, and standardized by these softwares[9].

However, AI models to predict patient clinical outcomes based on pre-procedural clinical and imaging information is limited. Our AI based machine learning toolkit helps predict functional outcomes beforehand and aid decision-making for mechanical thrombectomy (MT) based on readily available pre-procedural patient variables. Our model is unique as it uses two-stage machine learning approach – the first stage predicts patient's probability of death undergoing MT and the second stage predicts good or bad survival in a patient with low probability of death. This filtration method improves the accuracy of prediction to 80% which is superior when compared with previous models[5,9] and to the best of our knowledge there is no other predictive model which uses a two-stage machine learning approach.

Our study population has the highest number of patients when compared to the previously reported studies by using the standard statistical model and machine learning models. This study also used more prognostic factors which are

correlated with the clinical outcome in patients undergoing MT [5,9].

We plan to develop a web-based platform and a smart phone-based application where physicians can input patient data and obtain immediate prediction results in patients considered for MT treatment. This has the potential to globally improve the stroke care by helping the clinical teams in decision-making and selecting appropriate patient for MT.

CONCLUSIONS

The machine learning model with multiple standard pre-treatment clinical variables has the potential to improve the prediction of the outcome of LVO patients who receive MT. The tool is clinically useful as an aid to identify individual patients who may benefit from intra-arterial treatment for acute ischaemic stroke. The machine learning model was found to be superior to the standard statistical model based on logistic regression. The model has the probability to predict the clinical outcome of LVO patients better than the previously developed pre-treatment scoring methods. To the best of our knowledge, this is the first study to develop a two-stage machine learning model using the largest dataset from a wide range of stroke centres across many countries. Further validation would maximise its utility in routine clinical practice.

Conflict of interest
Dr Sanjeev Nayak is on the Advisory Board of Medtronic. Mr Qian Zhao is associated with DataRobot Inc. The affiliations did not relate to the research activity.

Ethics Declaration:
All human and animal studies have been approved by the appropriate ethics committee and have therefore been performed in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki and its later amendments.

REFERENCES

1. WHO Factsheet: The top 10 causes of death. Updated January 2017. Available at: 281 http://www.who.int/mediacentre/factsheets/fs310/en/ Accessed September 13 2020. 282

2. National Stroke Strategy. London: Department of Health 2001. Available at: 283 https://webarchive.nationalarchives.gov.uk/20130104224925/http://www.dh.gov.uk/prod_consum 284 _dh/groups/dh_digitalassets/documents/digitalasset/dh_081059.pdf Accessed September 13 2020.

3. Smith WS, Lev MH, English JD, et al. Significance of large vessel intracranial occlusion causing acute ischemic stroke and TIA. Stroke 2009;40:3834-40. doi:10.1161/STROKEAHA.109.561787.

4 Lima FO, Furie KL, Silva GS, et al. Prognosis of untreated strokes due to anterior circulation proximal intracranial arterial occlusions detected by use of computed tomography angiography. JAMA Neurol 2014;71:151-7. doi:10.1001/jamaneurol.2013.5007

5. Venema E, Mulder MJHL, Roozenbeek B, et al. Selection of patients for intra-arterial treatment for acute ischaemic stroke: development and validation of a clinical decision tool in two randomised trials. BMJ. 2017 May 3;357:j1710. doi: 10.1136/bmj.j1710. PMID: 28468840; PMCID: PMC5418887.

6. Powers WJ, Rabinstein AA, Ackerson T, Adeoye OM, Bambakidis NC, Becker K, et al.; American Heart Association Stroke Council. 2018 Guidelines for the Early Management of Patients With Acute Ischemic Stroke: a guideline for healthcare professionals from the American Heart Association/American Stroke Association.**Stroke.** 2018; 49:e46–e110. doi: 10.1161/STR.0000000000000158LinkGoogle Scholar.

7. Nogueira RG, Ribó M. Endovascular Treatment of Acute Stroke. Stroke. 2019 Sep;50(9):2612-2618. doi: 10.1161/STROKEAHA.119.023811. Epub 2019 Jul 25. PMID: 31340728.

8. Raseta M, Bazarova A, Wright H, Parrott A, Nayak S. A novel toolkit for the prediction of clinical outcomes following mechanical thrombectomy. Clin Radiol. 2020 Oct;75(10):795.e15-795.e21. doi: 10.1016/j.crad.2020.06.026. Epub 2020 Jul 24. PMID: 32718742.

9. Nishi H, Oishi N, Ishii A, et al. Predicting Clinical Outcomes of Large Vessel Occlusion Before Mechanical Thrombectomy Using Machine Learning. Stroke. 2019 Sep;50(9):2379-2388. doi: 10.1161/STROKEAHA.119.025411. Epub 2019 Aug 14. PMID: 31409267.

10. Mulder MJHL, Venema E, Roozenbeek B, et al. Towards personalised intra-arterial treatment of patients with acute ischaemic stroke: a study protocol for development and validation of a clinical decision aid. BMJ Open 2017;7:e013699.

11. DataRobot [2019/3/14]. Available from: https://www.datarobot.com/.

12. Kang J, Schwartz R, Flickinger J, Beriwal S. Machine Learning Approaches for Predicting Radiation Therapy Outcomes: A Clinician's Perspective. Int J Radiat Oncol Biol Phys. 2015; 93:1127–1135. https://doi.org/10.1016/j.ijrobp.2015.07.2286 PMID: 26581149

13. Brooks Carthon JM, Jarrin O, Sloane D, Kutney-Lee A. Variations in postoperative complications according to race, ethnicity, and sex in older adults. J Am Geriatr Soc. 2013; 61:1499–1507. https://doi. org/10.1111/jgs.12419 PMID: 24006851

14. Breiman L. Random Forests. Machine Learning. 2001; 45:5–32.

15. Friedman J. Greedy boosting approximation: a gradient boosting machine. Ann Stat. 2001; 29:1189–1232.

16. Craven GT, Lubbers N, Barros K, Tretiak S. Machine learning approaches for structural and thermodynamic properties of a Lennard-Jones fluid. J Chem Phys. 2020 Sep 14;153(10):104502. doi: 10.1063/5.0017894. PMID: 32933279.

17. Sutcliffe A, "Assessing the reliability of heuristic evaluation for Web site attractiveness and usability," Proceedings of the 35th Annual Hawaii International Conference on System Sciences, 2002, pp. 1838-1847, doi: 10.1109/HICSS.2002.994098.

18. Straka M, Albers GW, Bammer R. Real-time diffusion-perfusion mismatch analysis in acute stroke. J Magn Reson Imaging. 2010;32:1024–1037. doi: 10.1002/jmri.22338 30.

19. Campbell BC, Christensen S, Levi CR, Desmond PM, Donnan GA, Davis SM, et al. Cerebral blood flow is the optimal CT perfusion parameter for assessing infarct core. Stroke. 2011;42:3435–3440. doi: 10.1161/STROKEAHA.111.618355