A new machine learning algorithm with high interpretability for improving the safety and efficiency of thrombolysis for stroke patients: A hospital-based pilot study

Huiling Shao[1] (iD), Wing Chi Lawrence Chan[1], Heng Du[1],
Xiangyan Fiona Chen[1], Qilin Ma[2] and Zhiyu Shao[2]

Abstract

Background: Thrombolysis is the first-line treatment for patients with acute ischemic stroke. Previous studies leveraged machine learning to assist neurologists in selecting patients who could benefit the most from thrombolysis. However, when designing the algorithm, most of the previous algorithms traded interpretability for predictive power, making the algorithms hard to be trusted by neurologists and be used in real clinical practice.

Methods: Our proposed algorithm is an advanced version of classical k-nearest neighbors classification algorithm (KNN). We achieved high interpretability by changing the isotropy in feature space of classical KNN. We leveraged a cohort of 189 patients to prove that our algorithm maintains the interpretability of previous models while in the meantime improving the predictive power when compared with the existing algorithms. The predictive powers of models were assessed by area under the receiver operating characteristic curve (AUC).

Results: In terms of interpretability, only onset time, diabetes, and baseline National Institutes of Health Stroke Scale (NIHSS) were statistically significant and their contributions to the final prediction were forced to be proportional to their feature importance values by the rescaling formula we defined. In terms of predictive power, our advanced KNN (AUC 0.88) outperformed the classical KNN (AUC 0.75, $p = 0.0192$).

Conclusions: Our preliminary results show that the advanced KNN achieved high AUC and identified consistent significant clinical features as previous clinical trials/observational studies did. This model shows the potential to assist in thrombolysis patient selection for improving the successful rate of thrombolysis.

Keywords

Ischemic stroke, translational medicine, neuroimaging, machine learning, decision support systems

Background

Stroke is the third leading cause of death and chronic disability globally.[1] As a leading cause of adult disability, up to 74% of stroke survivors are dependent on activities of daily living,[2] which causes a huge burden to the society. Among different types of strokes, ischemic stroke is the most common, accounting for 87% compared to hemorrhagic stroke.[3] In the treatment of ischemic stroke, thrombolysis is the first-line treatment.[4] For acute ischemic stroke patients, a prompt treatment with thrombolytic drugs could restore blood flow before major brain damage has occurred and greatly improve short-term and long-term recovery after stroke,[5] as a result largely reducing the burden stroke brings to the society.

[1]Department of Health Technology and Informatics, The Hong Kong Polytechnic University, Kowloon, Hong Kong
[2]Department of Neurology, The First Affiliated Hospital of Xiamen University, Xiamen, China

Corresponding author:
Huiling Shao, Department of Health Technology and Informatics, The Hong Kong Polytechnic University, Room Y934, 9/F, Lee Shau Kee Building, Hung Hom, Kowloon, 999077, Hong Kong.
Email: hui-ling.shao@connect.polyu.hk

In most cases, thrombolysis therapy is subject to the latest guidelines. The guidelines are drawn up based on the large quantities of clinical evidence, therefore, the proposed eligibility and dosage consideration for thrombolysis treatment should normally be safe and efficient for most of the patients. However, in clinical practice, still, a number of patients present unpredictable outcomes after thrombolysis treatment, including symptomatic hemorrhage (13% among patients receiving thrombolysis)[6] and failed recanalization (37% among patients receiving thrombolysis),[7] suggesting that a more accurate patient-tailored clinical decision support tool based on guidelines to improve thrombolysis safety and efficiency is needed.

Previous studies leveraged machine learning models to assist neurologists in deciding the safety and efficiency for each patient more accurately[8–36]: they all simply reused the existent machine learning algorithm and trained the algorithm based on their patient cohort to predict thrombolysis outcome. However, when reusing the current machine learning algorithms, there is always a trade-off between flexibility and interpretability[37]: Inflexible algorithms have a restrictive ability to estimate the boundaries between different outcome classes, therefore presenting lower predictive power. But inflexible algorithms are often easy to be interpreted. On the other hand, flexible algorithms generate more accurate predicted outcome but suffer from low interpretability. If we are only interested in predictive power and the interpretability of the predictive model is simply not of interest, more flexible algorithms would be a good choice. However, in some settings, if the importance of inference outweighs the predictive power, we should turn to a more restrictive method since it's quite easy to tell which predictors are the causes of algorithm decision or result.

In our case, to make the model be trusted by clinicians and eligible to assist in clinical practice, the interpretability of algorithm is critical since all clinical decision support tools must go through clinical validation to be approved by the local authority before being used in real clinical practice. The interpretability of the algorithm allows to tell which predictors the algorithm leverages as important factors to be considered when predicting the clinical outcome or deciding the thrombolysis eligibility. These predictors then need to be confirmed related to the clinical outcome of patients going through thrombolysis by previous clinical trials or following clinical trials in case that the algorithm generates new features during training. For instance, according to the current clinical evidence, stroke patients with a small infarct core but a large penumbra will benefit a lot from thrombolysis with minimal side effects.[38] As a result, a clinically significant algorithm should not only predict thrombolysis outcome with high discriminative ability, but also devotes more attention to the penumbra related features.

To further improve the performance of current algorithms and offer the new algorithm the possibility to be used in real clinical scenario, in this research article, we propose a newly developed algorithm that combines the interpretability of restrictive algorithm and the predictive power of flexible algorithm. Our preliminary results show that the new algorithm maintains the interpretability while improving predictive power when compared with other common machine learning algorithms.

The research article is organized as follows: In the Related works section, we summarize previous thrombolysis outcome prediction models in terms of algorithm flexibility and interpretability, compare previous works with our proposed algorithm, and illustrate the originality of our study. In the Methods section, we introduce in detail our algorithm design and the methodology we used to demonstrate that the algorithm can maintain the interpretability while improving predictive power. In the Results section, we show the results of exploratory data analysis and model validation. In the Discussion section, we interpret the results of model validation, compare our findings with early studies, explain the clinical significance, and discuss limitations as well as propose future directions for the study.

Related works

Of the many algorithms used by previous thrombolysis outcome prediction studies, some can generate more flexible decision boundaries, others can only generate restrictive decision boundaries. Table 1 in the Supplemental Material summarizes previous thrombolysis outcome prediction studies in terms of algorithm flexibility and interpretability. The algorithms used by previous research, from the most restrictive to the most flexible, were logistic regression,[8,19–21,24,25,33,36] naïve Bayes classifier,[33] risk score,[9,10,12–16,35] nomogram,[22,23,28,30,31,34] tree-based machine learning models,[29,33,36] support vector machine,[11,17,29] and deep learning neural network.[11,26,27,29,32,33]

Most of the previous studies preferred restrictive models (logistic regression, naïve Bayes classifier, risk score, nomogram, and tree-based machine learning models). For these restrictive models, feature importance can be inferred, respectively, through feature coefficients, difference of the feature likelihood between the two classes, point assignment of each feature, graphic preliminary score assigned to each feature and weight metric.

Regarding flexible algorithms, there are two common ways to increase the interpretability: (1) A reactive approach to calculate individual predictor importance using SHapley Additive exPlanations (SHAP) framework proposed by Lundberg and Lee in 2017.[39] (2) A proactive approach to increase model predictive power by boosting interpretability, where a very popular example is the attention mechanism introduced in 2014[40] to allow the deep learning neural network decoder to leverage the most relevant parts of the input vectors in a flexible manner. Most of

the previous research adopting support vector machine and deep learning neural network didn't infer feature importance. A research in 2012[11] leveraged the reactive approach to calculate the individual predictor importance from neural networks and support vector machines. The proactive approach is preferred when compared with the reactive approach since the former can not only boost interpretability but increase model predictive power at the same time.

In this research article, we adopted the proactive approach to increase the interpretability of k-nearest neighbors classification algorithm (KNN), a rather flexible machine learning model with nonlinear and non-smooth boundaries based on the local geometry of the distribution of the data on the feature hyperplane. Our proposed algorithm can not only predict the thrombolysis therapy outcome with high discriminative ability but also tell which clinical features are the key to a good outcome.

Methods

Algorithm design

Our algorithm is an advanced version of KNN. We here changed the isotropy in feature space of classical KNN and developed a new algorithm that maintains the interpretability of previous models while in the meantime improving the predictive power when compared with the existing algorithms.

KNN is an intuitive supervised machine learning algorithm where an object is classified by a plurality vote of its K neighbors, with K a positive integer calculated in the training process.[41] When implementing KNN, the first step is to transform each data point into a vector composed of the mathematical values of each of its features. The algorithm computes the distance between each data point (vector) and then finds the K-nearest points to the test data. The test data will be assigned to the class most common among its K-nearest neighbors. KNN is widely used in medical fields because the algorithm simulates best the evidence-based working flow of a clinician: the K-nearest points correspond to the K patients with the most similar clinical profiles to the new patient. However, a significant shortcoming of KNN is the isotropy in feature space and no interpretability: in classical logistic regression, for example, the contribution of each feature to the classification result can be easily known by examining the statistical significance and comparing the coefficient of each feature in the model. In KNN setting, the algorithm often uses the Euclidean distance between two data points to find the nearest neighbors. Euclidean distance is sensitive to the standard deviation (SD). The features with high SD will weigh more than features with low SD. Therefore, after standardization of data preprocessing, each feature will follow a normal distribution with mean at 0 and SD of 1 and is forced to contribute equally to the prediction result,

which might compromise the predictive power since each feature has its own importance to predict the output. Besides, there is no way to infer significant features. To solve this problem, we developed an advanced KNN algorithm by introducing the inference power into classical KNN algorithm: The pseudocode of training and testing algorithms of the advanced KNN model can be found, respectively, in Algorithm 1 and Algorithm 2 in the Supplemental material.

In training phase, the following parameters need to be defined in algorithm input: the p-value threshold p to determine significant features, the set of hyperparameters M_t, M_l to calculate the feature importance, and the number of nearest points k to predict.

Let X_train denote the training data features matrix and y_train denotes the training data outcome array. A logistic regression was first built on $\{X\_train, y\_train\}$. We used the z-statistic to test if coefficient is null and calculate the p-value for each coefficient. Let P denote the set of p-value for each coefficient. We defined features whose associated coefficient p-value is inferior to a threshold p as statistically significant features. Let S denote the set of indexes of statistically significant features in X_train.

Then the feature importance of each statistically significant feature in X_train was calculated. Feature importance calculation refers to techniques that assign a score to input features based on how useful they are at predicting a target variable. Popular techniques include coefficient calculated as part of linear models and weight metric of decision trees. When building a linear model, the coefficients describe the mathematical relationship between each independent variable and the dependent variable. When building a decision tree model, weight metric of a feature is the number of times a feature is used to split the data across all trees. Let $M\_t = \{activate\_t, n\_estimators, h\}$ denote the set of hyperparameters of decision tress technique to calculate feature importance, where $activate\_t = 1$ when decision tree is chosen to calculate feature importance and $activate\_t = 0$ if not; $n\_estimators$ is the number of trees in the forest when bagging is used to avoid overfitting; h is the maximum depth of trees in the forest. Let $M\_l = \{activate\_l\}$ denote the set of hyperparameters of linear models technique to calculate feature importance, where $activate\_l = 1$ when linear models is chosen to calculate feature importance and $activate\_l = 0$ if not. We also defined F as the set of feature importance value of each significant feature in X_train.

Next SD along the feature axis for each significant feature in X_train was adjusted by the following formula: $x^* = \frac{x-\mu}{\sigma} \times f$, where x^* is the rescaled feature value after SD adjust, x is the original feature value, μ is the mean, σ is the original deviation and f is the corresponding feature importance value.

The rescaled significant features are the input of the following KNN. In the KNN model training, we found the k-nearest points to make the final prediction.

Our advanced KNN algorithm improves greatly the interpretability of classical KNN: the z-statistic and associated p-value allow to tell the level of certainty of the statistical significance of the input features used by advanced KNN algorithm to make final prediction. Since that original feature SD had been rescaled based on their feature importance and that in KNN setting the features with high SD will weigh more than features with low SD, we make sure that in the advanced KNN algorithm, features with high importance value will make more contributions to predict the outcome. In conclusion, we ensured that in our advanced KNN algorithm, it's the statistically significant and important variables who are taking effects and each variable makes a proper contribution to the final prediction result, which greatly improves the robustness of the model and decreases the overfitting possibility.

In testing phase, we first included features only statistically significant in training dataset, then rescaled the test data using the same formula, assuming that the test data share the same mean, SD, and importance value with the training data, and finally made prediction based on the nearest k points.

In Algorithm 1 and Algorithm 2 as available in the Supplemental Material, we also annotated each step with time complexity. In training algorithm, $max\_1 = \max$ $(O(n\_estimators \times n \times \log n \times d), O(n\_estimators \times (2^{h+1} - 1)))$ denotes the total time complexity of using decision trees to calculate feature importance. $max\_2 = \max(O(nd), O(d))$ denotes the total time complexity of using linear models to calculate feature importance. Since only one method would be used to calculate feature importance, $max\_3 = \max(max\_1, max\_2)$ denotes the total time complexity of feature importance calculation. The time complexity of training algorithm would be $max\_train = \max(O(knd), O(n\_estimators \times n \times \log n \times d), O(n\_estimators \times (2^{h+1} - 1)))$. The time complexity of test algorithm would be $max\_test = \max(O(knd), O(n \times \log n))$.

Participants and clinical assessments

We leveraged a cohort with acute ischemic stroke undergoing thrombolysis, but without following endovascular thrombectomy, to prove that our new algorithm maintains the interpretability of previous models while in the meantime improving the predictive power when compared with existent algorithms. The data of cohort were collected from The First Affiliated Hospital of Xiamen University from November 2013 to May 2020. The patient consent is waived as the retrospective non-identifiable data were used. The approval number is SL-2020KY023-01.

The inclusion criteria were as follows: (1) diagnosis of early ischemic stroke; (2) thrombolysis treatment with intravenous recombinant tissue plasminogen activator; (3) age superior to 18; (4) completion of baseline National Institutes of Health Stroke Scale (NIHSS) assessments; and (5) completion of 3-month modified Rankin Scale (mRS). The exclusion criterion was endovascular thrombectomy treatment following thrombolysis treatment.

The following information was obtained from each patient's clinical profile, namely, general clinical information including age and sex, risk factors including cardioembolic risk factors, diabetes, hyperlipidemia and hypertension, baseline information including body mass index (BMI), onset time and baseline NIHSS. A patient was considered to have cardioembolic risk factors if chronic rheumatic heart disease (International Classification of Disease 10th Revision (ICD-10): I05-I09), or nonrheumatic mitral valve disorders (ICD-10: I34), or nonrheumatic aortic valve disorders (ICD-10: I35), or endocarditis, valve unspecified (ICD-10: I38), or atrial fibrillation and flutter (ICD-10: I48) were found in medical history.

Additionally, the thrombolysis dosage was recorded for treatment option definition and 3-month mRS was recorded for clinical outcome assessment definition. Two thrombolysis treatment options were available: a thrombolysis dosage $<0.6mg/kg$ was defined as low dosage (LD), a dosage $\geq 0.6mg/kg$ was defined as normal dosage (ND). A favorable outcome (FO) was defined as mRS ≤ 2, an unfavorable outcome (UFO) was defined as mRS >2.

Statistical analysis

Analysis of variance (ANOVA) test with no assumption of equal variances and Fisher's exact test were employed for statistical analysis on characteristics of patients at baseline in different groups (LD vs. ND; FO vs. UFO). A logistic regression with z-statistic test was employed to test the statistical correlation between treatment options and the other clinical features. Treatment option shall be excluded from the input dataset if high correlation is found.

Validation of the predictive models

The proposed algorithm was trained on training dataset and validated on the test dataset. Since the proposed model is an advanced version of classical KNN, the performance of proposed advanced KNN was tested in comparison with the performance of classical KNN. We also varied feature importance calculation hyperparameters to test their impact on algorithm discriminative ability while all other hyperparameters were held constant. The predictive power was assessed according to area under the receiver operating characteristic curve (AUC). Pairwise AUC comparisons between models were tested using bootstrap

method, with 1000 bootstrap replicates.[42] Other classification metrics such as accuracy, precision, recall, and F1 score were also reported for reference.

Results

Exploratory data analysis

Initially 202 patients met the inclusion criteria, and 13 patients were excluded due to endovascular thrombectomy treatment following thrombolysis treatment. Finally, a total of 189 patients were included in the final cohort. The dataset was split into a training dataset (132 patients) and a test dataset (57 patients).

Before constructing the outcome prediction model, statistical correlation between treatment option and the other clinical features was first tested. Table 1 shows, respectively, characteristics of patients with LD and ND at baseline in training and test datasets. In training dataset, cardioembolic risk factors and baseline NIHSS were significantly different between LD and ND groups. In test dataset, significant differences were only evident for BMI. We then constructed a logistic regression on

training dataset to test the correlation between treatment option and the other clinical features. The estimated coefficients and associated p-value are shown in Table 2. According to the z-statistic test result, the thrombolysis treatment option was correlated with sex, cardioembolic risk factors, hypertension, and baseline NIHSS in training dataset, meaning that these significant variables were factors considered by clinicians when making thrombolysis treatment option. Therefore, when building the outcome prediction model, treatment option should not be included as an input variable given the high correlation. Table 3 shows, respectively, characteristics of the patient with FO and patients with UFO at baseline in training and test datasets. In training dataset, sex, age, cardioembolic risk factors, diabetes, onset time, and baseline NIHSS were significantly different between FO and UFO groups. In test dataset, significant differences were only evident for baseline NIHSS score.

Model development and validation

The advanced KNN prediction model with the following parameters proved to have the optimal discriminative ability

Table 1. Characteristics of patients with low dosage (LD) and normal dosage (ND) at baseline in training and test datasets.

	Training dataset			Test dataset		
	LD (N = 36)	ND (N = 96)	p-value	LD (N = 14)	ND (N = 43)	p-value
General clinical information						
Sex (male)	53.13%	65.63%	0.1057	64.29%	65.12%	1
Age (years), median(range)	66 (45–86)	66 (22–87)	0.3297	64 (40–88)	66 (31–88)	0.3163
Risk factors						
Cardioembolic risk factors	15.63%	31.25%	0.01649	42.86%	39.53%	1
Diabetes	23.96%	20.83%	0.7295	14.29%	16.28%	1
Hyperlipidemia	5.21%	12.50%	0.1254	0%	18.60%	0.1792
Hypertension	59.38%	68.75%	0.2288	50%	53.49%	1
Baseline information						
BMI	24.37 ± 3.47	23.54 ± 2.79	0.06797	21.93 ± 3.28	24.02 ± 2.99	0.04689
Onset time (hours)	2.76 ± 0.98	2.61 ± 0.91	0.2695	2.93 ± 1.31	2.64 ± 1.43	0.4912
NIHSS, median (range)	7 (2–22)	10 (1–23)	0.03013	11 (2–30)	12 (1–29)	0.8885

Note: Data are presented as mean ± SD or n(%). Clinical feature is considered significantly different in two groups LD and ND when its associated p-value ≤ 0.05.
BMI: body mass index; NIHSS: National Institutes of Health Stroke Scale.

Table 2. The coefficient and associated *p*-value of each clinical feature in the logistic regression model to test the statistical correlation between treatment option and the other clinical features.

	Coefficient	*p*-value
Sex	0.7671	0.0207
Age	−0.0104	0.3048
Cardioembolic risk factors	0.9093	0.0266
Diabetes	−0.4451	0.2592
Hyperlipidemia	1.0713	0.0661
Hypertension	0.7653	0.0267
BMI	−0.0377	0.2366
Onset time	−0.0823	0.6171
Baseline NIHSS	0.0643	0.0363

Note: Clinical feature is considered significant when its associated *p*-value ≤ 0.05.
BMI: body mass index; NIHSS: National Institutes of Health Stroke Scale.

on our test dataset with an AUC of 0.88: the *p*-value threshold 0.1, the weight metric of decision trees to calculate the feature importance with $n\_estimators = 100$ and h $= 4$, the number of nearest points 13 to predict. Three variables, onset time, diabetes, and baseline NIHSS, were retained in the significant feature selection with a *p*-value, respectively, of 0.0316, 0.0806, and 0.0000. The feature importance of onset time, diabetes, and baseline NIHSS score were, respectively, 2.0, 4.0, and 24.0 based on the weight metric of decision trees.

Figure 1 shows that on test dataset, our advanced KNN outperformed the classical KNN in terms of AUC. And with AUC comparison test, there were significant differences between classical KNN and advanced KNN ($p = 0.0192$) on test dataset. The comparison of other classification metrics of advanced KNN and classical KNN can be found in Table 4.

Based on the optimal model, feature importance calculation hyperparameters were varied to test their impact on algorithm performance while all other hyperparameters were held constant. The feature importance of onset time, diabetes, and baseline NIHSS score were respectively 0.3904, 0.7738 and 0.1715 based on the coefficients of

Table 3. Characteristics of patients with favorable outcome (FO) and unfavorable outcome (UFO) at baseline in training and test datasets.

	Training dataset			Test dataset		
	FO (N = 90)	UFO (N = 42)	*p*-value	FO (N = 36)	UFO (N = 21)	*p*-value
General clinical information						
Sex (male)	70.00%	51.11%	0.01441	69.44%	57.14%	0.3973
Age (years), median (range)	65 (22–87)	67 (42–86)	0.01832	64 (31–83)	68 (40–88)	0.1871
Risk factors						
Cardioembolic risk factors	22.22%	40.00%	0.01534	33.33%	52.38%	0.1751
Diabetes	28.89%	13.33%	0.01684	16.67%	14.29%	1
Hyperlipidemia	13.33%	5.56%	0.1242	11.11%	19.05%	0.449
Hypertension	70.00%	68.89%	1	52.78%	52.38%	1
Baseline information						
BMI	23.78 ± 3.16	24.04 ± 2.41	0.5326	23.43 ± 3.02	23.63 ± 3.46	0.8326
Onset time (hours)	2.73 ± 1.01	2.36 ± 0.93	0.01267	2.63 ± 1.02	2.86 ± 1.90	0.6093
NIHSS, median (range)	7 (1–20)	13 (4–23)	2.522×10^{-10}	6 (1–30)	17 (11–29)	1.877×10^{-6}

Note: Data are presented as mean ± SD or n (%). Clinical feature is considered significantly different in two groups FO and UFO when its associated *p*-value ≤ 0.05.
BMI: body mass index; NIHSS: National Institutes of Health Stroke Scale.
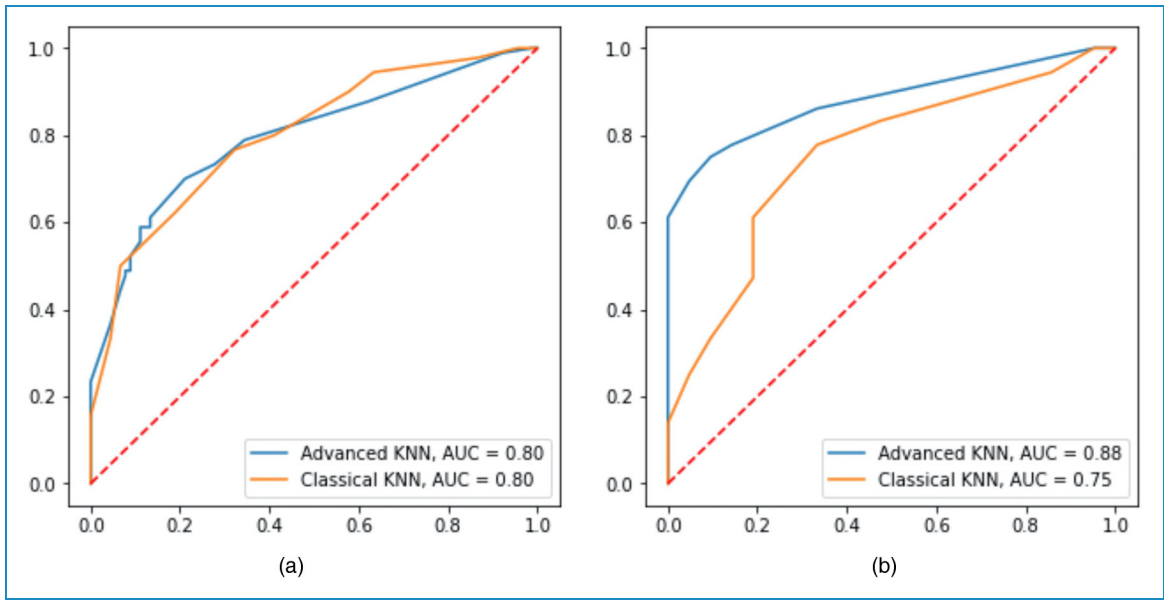
Figure 1. Comparison of classical k-nearest neighbors classification algorithm (KNN) and advanced KNN. The left figure (a) shows the AUC obtained on training dataset, using respectively classical KNN (orange) and advanced KNN (blue). The right figure (b) shows the area under the receiver operating characteristic curve (AUC) obtained on test dataset, using respectively classical KNN (orange) and advanced KNN (blue).

Table 4. Performance comparison on test dataset of advanced k-nearest neighbors classification algorithm (KNN) with weight metric of decision trees, classical KNN, and advanced KNN with coefficient of linear model in terms of accuracy, precision, recall, and F1 score.

Advanced KNN with weight metric of decision trees				
	Accuracy	Precision	Recall	F1 score
Unfavorable outcome (UFO)	0.75	0.6	1	0.75
Favorable outcome (FO)		1	0.61	0.76
Classical KNN				
	Accuracy	Precision	Recall	F1 score
UFO	0.6	0.47	0.81	0.6
FO		0.81	0.47	0.6
Advanced KNN with coefficient of linear model				
	Accuracy	Precision	Recall	F1 score
UFO	0.72	0.59	0.76	0.67
FO		0.83	0.69	0.76

linear model. Figure 2 shows that given the same p-value threshold and the same number of nearest points, how the method to calculate the feature importance affected the final prediction power of the model. On test dataset, the AUC of optimal model with weight metric as feature importance calculation method 0.88 was superior to the AUC of the model with coefficient as feature importance calculation method 0.75. With AUC comparison test, there were
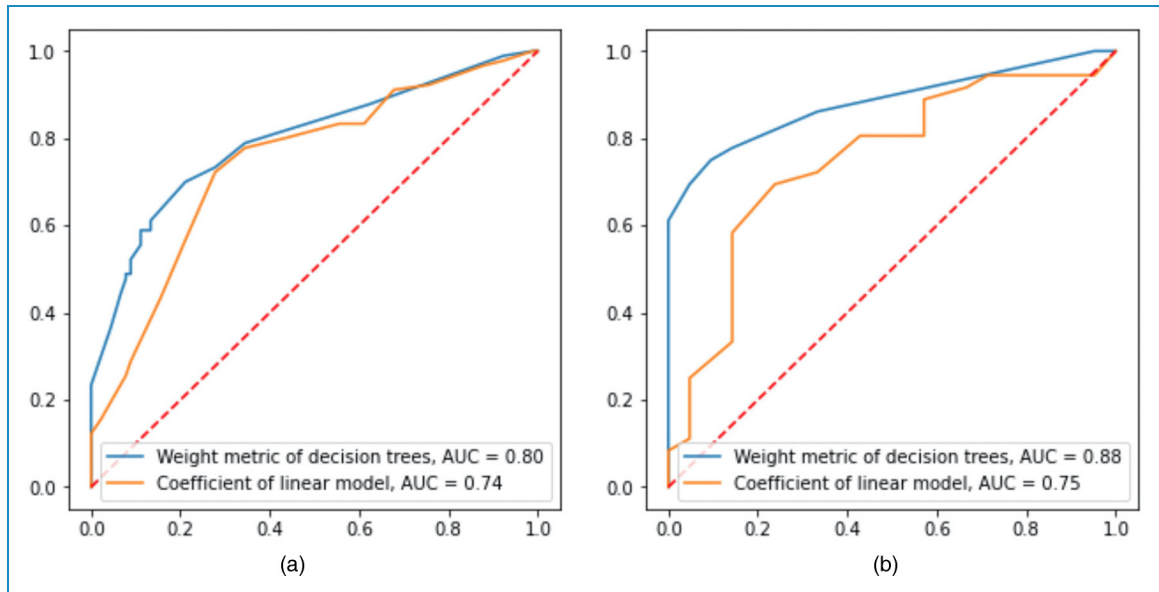
Figure 2. The impact of feature importance calculation hyperparameters variation on algorithm predictive power. The left figure (a) shows the area under the receiver operating characteristic curve (AUC) obtained on training dataset, using respectively the weight metric (blue) and the linear coefficient (orange) as feature importance calculation method. The right figure (b) shows the AUC obtained on test dataset, using respectively the weight metric (blue) and the linear coefficient (orange) as feature importance calculation method.

significant differences between two models with different feature importance calculation methods ($p = 0.00048$). The comparison of other classification metrics of advanced KNN with coefficient of linear models and advanced KNN with weight metric of decision trees can be found in Table 4.

Discussion

Summary of main findings

In this article, we proposed a new algorithm that combines the interpretability of restrictive algorithm and the predictive power of flexible algorithm for thrombolysis outcome prediction. Our proposed algorithm is an advanced version of classical KNN. We achieved high interpretability by changing the isotropy in feature space of classical KNN.

Our preliminary results show that our advanced KNN maintains high interpretability while not compromising the predictive power: model inference revealed that three variables: onset time, diabetes, and baseline NIHSS proved significant feature importance in outcome prediction. Compared with the classical KNN, our advanced KNN outperformed the classical KNN in terms of AUC (0.88 vs. 0.75, $p = 0.0192$).

Moreover, the predictive power of algorithm is highly dependent on the feature importance values: based on the optimal model, feature importance calculation hyperparameters were replaced with those of linear models to test their impact on algorithm discriminative ability while all other hyperparameters were held constant. There were

significant differences between the two models with different feature importance calculation methods (AUC 0.88 vs. 0.75, $p = 0.00048$), indicating that the weight metric of decision trees outperformed coefficient of linear model in terms of calculation of the contribution of each variable to the final prediction result.

Comparison with previous studies

As we have stated in the Related works section , most of previous thrombolysis outcome prediction studies preferred restrictive algorithms (logistic regression, naïve Bayes classifier, risk score, nomogram, and tree-based machine learning models) for the high interpretability, while compromising the ability to generate true decision boundaries. Most of the previous research adopting flexible algorithms (support vector machine and deep learning neural network) didn't infer feature importance, except for the one published in 2012[11] which leveraged the reactive approach to calculate individual predictor importance from neural networks and support vector machine.

In this research article, we adopted a proactive approach to increase the interpretability of KNN, a rather flexible machine learning algorithm, by changing the isotropy in feature space while in the meantime improving its predictive power. The results indicated that our advanced KNN outperformed classical KNN in terms of discriminative ability. Besides, model inference revealed that three predictors, onset time, diabetes, and baseline NIHSS proved

significant feature importance in thrombolysis outcome prediction.

Consistent with early clinical observational studies and clinical trials, onset time, diabetes, and baseline NIHSS were leveraged as important predictors by our algorithm when predicting the thrombolysis outcome: The phrase "time is brain" emphasizes that as stroke progresses, human nerve tissue is rapidly lost, requiring emergent therapy.[43] Research also reveals that early treatment with intravenous alteplase improves outcome.[44–46] History of diabetes mellitus and admission glucose level (AGL) is associated with poor clinical outcomes after thrombolysis since they are related to lower recanalization rates,[47,48] indicating an impaired fibrinolytic response in the setting of elevated blood glucose concentration.[49] Both chronic and acute hyperglycemia contribute to coagulation activation,[50,51] whereas hyperinsulinemia decreases fibrinolytic activity by increasing plasminogen activator inhibitor production.[52,53] High baseline NIHSS, which in most cases represent severe or diffuse neuron impairment due to ischemic stroke, has also been demonstrated to be related to poor outcome[54,55] by multiple studies.

Clinical implications

In the literature review[56] we published earlier, we reviewed the literature on the feasibility of machine learning models to assist in stroke thrombolysis. We identified two factors that will hinder the implementation of models in the thrombolysis setting: the interpretability and the processing time of the model. This research study aims to address the issues related to interpretability. The interpretability of the algorithm allows neurologists to better understand the decision making process of an AI algorithm, which improves the trust between neurologists and machines.

Limitations and future directions

The limitation of this study is concerned with both the model interpretability and predictive power. First, the sample size was relatively small and was selected from a single medical center. Further external validations using multicenter data are recommended to validate the predictive power of the model. Second, no radiological features were included in this pilot study. Given the massive information the medical images contain regarding the thrombolysis outcome prediction,[57] we shall extract and include penumbra and vascular features from medical images as we proposed in our early study[56] to test if the algorithm can identify the same significant radiological features as previous clinical trials/observational studies did. The application of the model to a larger cohort from multicenter with radiological features included would generate more convincible evidence for the eligibility of our algorithm to assist in thrombolysis patient selection in clinical practice.

Conclusions

In summary, in this study, we proposed an advanced KNN which can maintain high interpretability while not compromising the predictive power when compared with classical KNN. The advanced KNN achieved high AUC and identified consistent significant clinical features as previous clinical trials/observational studies did. This model shows the potential to assist in thrombolysis patient selection for improving the successful rate of thrombolysis.

ORCID iD: Huiling Shao https://orcid.org/0000-0002-9115-8721

References

1. Feigin VL, Norrving B and Mensah GA. Global burden of stroke. *Circ Res* 2017; 120: 439–448.
2. Miller EL, Murray L, Richards L, et al. Comprehensive overview of nursing and interdisciplinary rehabilitation care of the stroke patient: a scientific statement from the American Heart Association. *Stroke* 2010; 41: 2402–2448.
3. Lloyd-Jones D, Adams R, Carnethon M, et al. Heart disease and stroke statistics–2009 update: a report from the American heart association statistics committee and stroke statistics subcommittee. *Circulation* 2009; 119: 480–486.

4. Jivan K, Ranchod K and Modi G. Management of ischaemic stroke in the acute setting: review of the current status. *Cardiovasc J Afr* 2013; 24: 86–92.

5. Wardlaw JM, Murray V, Berge E, et al. Thrombolysis for acute ischaemic stroke. *Cochrane Database Syst Rev* 2014; 2014: CD000213.

6. Miller DJ, Simpson JR and Silver B. Safety of thrombolysis in acute ischemic stroke: a review of complications, risk factors, and newer technologies. *Neurohospitalist* 2011; 1: 138–147.

7. Davalos A. Thrombolysis in acute ischemic stroke: successes, failures, and new hopes. *Cerebrovasc Dis* 2005; 20: 135–139.

8. Kent DM, Selker HP, Ruthazer R, et al. The stroke-thrombolytic predictive instrument: a predictive instrument for intravenous thrombolysis in acute ischemic stroke. *Stroke* 2006; 37: 2957–2962.

9. Cucchiara B, Tanne D, Levine SR, et al. A risk score to predict intracranial hemorrhage after recombinant tissue plasminogen activator for acute ischemic stroke. *J Stroke Cerebrovasc Dis* 2008; 17: 331–333.

10. Lou M, Safdar A, Mehdiratta M, et al. The HAT score: a simple grading scale for predicting hemorrhage after thrombolysis. *Neurology* 2008; 71: 1417–1423.

11. Dharmasaroja P and Dharmasaroja PA. Prediction of intracerebral hemorrhage following thrombolytic therapy for acute ischemic stroke using multiple artificial neural networks. *Neurol Res* 2012; 34: 120–128.

12. Mazya M, Egido JA, Ford GA, et al. Predicting the risk of symptomatic intracerebral hemorrhage in ischemic stroke treated with intravenous alteplase: safe implementation of treatments in stroke (SITS) symptomatic intracerebral hemorrhage risk score. *Stroke* 2012; 43: 1524–1531.

13. Menon BK, Saver JL, Prabhakaran S, et al. Risk score for intracranial hemorrhage in patients with acute ischemic stroke treated with intravenous tissue-type plasminogen activator. *Stroke* 2012; 43: 2293–2299.

14. Strbian D, Engelter S, Michel P, et al. Symptomatic intracranial hemorrhage after stroke thrombolysis: the SEDAN score. *Ann Neurol* 2012; 71: 634–641.

15. Strbian D, Meretoja A, Ahlhelm FJ, et al. Predicting outcome of IV thrombolysis-treated ischemic stroke patients: the DRAGON score. *Neurology* 2012; 78: 427–432.

16. Saposnik G, Guzik AK, Reeves M, et al. Stroke prognostication using age and NIH stroke scale: SPAN-100. *Neurology* 2013; 80: 21–28.

17. Bentley P, Ganesalingam J, Carlton Jones AL, et al. Prediction of stroke thrombolysis outcome using CT brain machine learning. *Neuroimage Clin* 2014; 4: 635–640.

18. Flint AC, Gupta R, Smith WS, et al. The THRIVE score predicts symptomatic intracerebral hemorrhage after intravenous tPA administration in SITS-MOST *Int J Stroke* 2014; 9: 705–710.

19. Kent DM, Ruthazer R, Decker C, et al. Development and validation of a simplified Stroke-Thrombolytic Predictive Instrument. *Neurology* 2015; 85: 94–949.

20. Lee JS, Kim CK, Kang J, et al. A novel computerized clinical decision support system for treating thrombolysis in patients with acute ischemic stroke. *J Stroke* 2015; 17: 199–209.

21. Lokeskrawee T, Muengtaweepongsa S, Patumanond J, et al. Prediction of symptomatic intracranial hemorrhage after intravenous thrombolysis in acute ischemic stroke: the symptomatic intracranial hemorrhage score. *J Stroke Cerebrovasc Dis* 2017; 26: 2622–2629.

22. Cappellari M, Turcato G, Forlivesi S, et al. The START nomogram for individualized prediction of the probability of unfavorable outcome after intravenous thrombolysis for stroke. *Int J Stroke* 2018; 13: 700–706.

23. Cappellari M, Turcato G, Forlivesi S, et al. STARTING-SICH nomogram to predict symptomatic intracerebral hemorrhage after intravenous thrombolysis for stroke. *Stroke* 2018; 49: 397–404.

24. Tang TY, Jiao Y, Cui Y, et al. Development and validation of a penumbra-based predictive model for thrombolysis outcome in acute ischemic stroke patients. *EBioMedicine* 2018; 35: 251–259.

25. Nisar T, Hanumanthu R and Khandelwal P. Symptomatic intracerebral hemorrhage after intravenous thrombolysis: predictive factors and validation of prediction models. *J Stroke Cerebrovasc Dis* 2019; 28: 104360.

26. Bacchi S, Zerner T, Oakden-Rayner L, et al. Deep learning in the prediction of ischaemic stroke thrombolysis functional outcomes: a pilot study. *Acad Radiol* 2020; 27: e19–e23.

27. Chung CC, Chan L, Bamodu OA, et al. Artificial neural network based prediction of postthrombolysis intracerebral hemorrhage and death. *Sci Rep* 2020; 10: 20501.

28. Lv S, Song Y, Zhang FL, et al. Early prediction of the 3-month outcome for individual acute ischemic stroke patients who received intravenous thrombolysis using the N2H3 nomogram model. *Ther Adv Neurol Disord* 2020; 13: 1756286420953054.

29. Wang F, Huang Y, Xia Y, et al. Personalized risk prediction of symptomatic intracerebral hemorrhage after stroke thrombolysis using a machine-learning model. *Ther Adv Neurol Disord* 2020; 13: 1756286420902358.

30. Wu Y, Chen H, Liu X, et al. A new nomogram for individualized prediction of the probability of hemorrhagic transformation after intravenous thrombolysis for ischemic stroke patients. *BMC Neurol* 2020; 20: 426.

31. Zhou Z, Yin X, Niu Q, et al. Risk factors and a nomogram for predicting intracranial hemorrhage in stroke patients undergoing thrombolysis. *Neuropsychiatr Dis Treat* 2020; 16: 1189–1197.

32. Chen Z, Li Q, Li R, et al. Ensemble learning accurately predicts the potential benefits of thrombolytic therapy in acute ischemic stroke. *Quant Imaging Med Surg* 2021; 11: 3978–3989.

33. Chung CC, Bamodu OA, Hong CT, et al. Application of machine learning-based models to boost the predictive power of the SPAN index. *Int J Neurosci* 2021 Feb; 3: 1–11. DOI: 10.1080/00207454.2021.1881092.

34. Huang L, Li F, Huang C, et al. A novel nomogram for predicting poor 6-month function in patients with acute ischemic stroke receiving thrombolysis. *J Cardiovasc Nurs* 2022; 37: E206–E216.

35. Soni M, Wijeratne T and Ackland DC. A risk score for prediction of symptomatic intracerebral haemorrhage following thrombolysis. *Int J Med Inform* 2021; 156: 104586.

36. Zhu B, Zhao J, Cao M, et al. Predicting 1-hour thrombolysis effect of r-tPA in patients with acute ischemic stroke using machine learning algorithm. *Front Pharmacol* 2021; 12: 759782.

37. James G, Witten D, Hastie T, et al. *An Introduction to Statistical Learning*. New York, NY: Springer, 2013.

38. Gravanis I and Tsirka SE. Tissue-type plasminogen activator as a therapeutic target in stroke. *Expert Opin Ther Targets* 2008; 12: 159–170.

39. Lundberg SM and Lee S-I. A unified approach to interpreting model predictions. *Adv Neural Inf Process Syst* 2017; 30: 1–10.

40. Bahdanau D, Cho K and Bengio Y. Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473, 2014

41. Altman NS. An Introduction to kernel and nearest-neighbor nonparametric regression. *Am Stat* 1992; 46: 175–185.

42. Robin X, Turck N, Hainard A, et al. pROC: an open-source package for R and S + to analyze and compare ROC curves. *BMC Bioinformatics* 2011; 12: 77.

43. Saver JL. Time is brain–quantified. *Stroke* 2006; 37: 263–266.

44. Lees KR, Bluhmki E, von Kummer R, et al. Time to treatment with intravenous alteplase and outcome in stroke: an updated pooled analysis of ECASS, ATLANTIS, NINDS, and EPITHET trials. *Lancet* 2010; 375: 1695–1703.

45. Strbian D, Ringleb P, Michel P, et al. Ultra-early intravenous stroke thrombolysis: do all patients benefit similarly? *Stroke* 2013; 44: 2913–2916.

46. Meretoja A, Keshtkaran M, Saver JL, et al. Stroke thrombolysis: save a minute, save a day. *Stroke* 2014; 45: 1053–1058.

47. Ribo M, Molina C, Montaner J, et al. Acute hyperglycemia state is associated with lower tPA-induced recanalization rates in stroke patients. *Stroke* 2005; 36: 1705–1709.

48. Zangerle A, Kiechl S, Spiegel M, et al. Recanalization after thrombolysis in stroke patients: predictors and prognostic implications. *Neurology* 2007; 68: 39–44.

49. Desilles JP, Meseguer E, Labreuche J, et al. Diabetes mellitus, admission glucose, and outcomes after stroke thrombolysis: a registry and systematic review. *Stroke* 2013; 44: 1915–1923.

50. Vaidyula VR, Rao AK, Mozzoli M, et al. Effects of hyperglycemia and hyperinsulinemia on circulating tissue factor procoagulant activity and platelet CD40 ligand. *Diabetes* 2006; 55: 202–208.

51. Lemkes BA, Hermanides J, Devries JH, et al. Hyperglycemia: a prothrombotic factor? *J Thromb Haemost* 2010; 8: 1663–1669.

52. Pandolfi A, Giaccari A, Cilli C, et al. Acute hyperglycemia and acute hyperinsulinemia decrease plasma fibrinolytic activity and increase plasminogen activator inhibitor type 1 in the rat. *Acta Diabetol* 2001; 38: 71–76.

53. Meigs JB, Mittleman MA, Nathan DM, et al. Hyperinsulinemia, hyperglycemia, and impaired hemostasis: the framingham offspring study. *JAMA* 2000; 283: 221–228.

54. Wahlgren N, Ahmed N, Eriksson N, et al. Multivariable analysis of outcome predictors and adjustment of main outcome results to baseline data profile in randomized controlled trials: safe implementation of thrombolysis in stroke-MOnitoring STudy (SITS-MOST). *Stroke* 2008; 39: 3316–3322.

55. Stefanovic Budimkic M, Pekmezovic T, Beslac-Bumbasirevic L, et al. Long-Term prognosis in ischemic stroke patients treated with intravenous thrombolytic therapy. *J Stroke Cerebrovasc Dis* 2017; 26: 196–203.

56. Shao H, Chen X, Ma Q, et al. The feasibility and accuracy of machine learning in improving safety and efficiency of thrombolysis for patients with stroke: Literature review and proposed improvements. *Front Neurol* 2022; 13: 934929. DOI: 10.3389/fneur.2022.934929.

57. El-Koussy M, Schroth G, Brekenfeld C, et al. Imaging of acute ischemic stroke. *Eur Neurol* 2014; 72: 309–316.