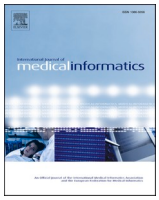
International Journal of Medical Informatics 152 (2021) 104505

Contents lists available at ScienceDirect

International Journal of Medical Informatics

journal homepage: www.elsevier.com/locate/ijmedinf



Developing a stroke alert trigger for clinical decision support at emergency triage using machine learning

Sheng-Feng Sung [a,b,c,1], Ling-Chien Hung [a,1], Ya-Han Hu [d,*]

[a] Division of Neurology, Department of Internal Medicine, Ditmanson Medical Foundation Chia-Yi Christian Hospital, Chiayi City, Taiwan
[b] Department of Information Management and Institute of Healthcare Information Management, National Chung Cheng University, Chiayi County, Taiwan
[c] Department of Nursing, Min-Hwei Junior College of Health Care Management, Tainan, Taiwan
[d] Department of Information Management, National Central University, Taoyuan City, Taiwan

ARTICLE INFO

Keywords:
Acute stroke
Clinical decision support
Emergency department
Triage

ABSTRACT

Background: Acute stroke is an urgent medical condition that requires immediate assessment and treatment. Prompt identification of patients with suspected stroke at emergency department (ED) triage followed by timely activation of code stroke systems is the key to successful management of stroke. While false negative detection of stroke may prevent patients from receiving optimal treatment, excessive false positive alarms will substantially burden stroke neurologists. This study aimed to develop a stroke-alert trigger to identify patients with suspected stroke at ED triage.

Methods: Patients who arrived at the ED within 12 h of symptom onset and were suspected of a stroke or transient ischemic attack or triaged with a stroke-related symptom were included. Clinical features at ED triage were collected, including the presenting complaint, triage level, self-reported medical history (hypertension, diabetes, hyperlipidemia, heart disease, and prior stroke), vital signs, and presence of atrial fibrillation. Three rule-based algorithms, ie, Face Arm Speech Test (FAST) and two flavors of Balance, Eyes, FAST (BE-FAST), and six machine learning (ML) techniques with various resampling methods were used to build classifiers for identification of patients with suspected stroke. Logistic regression (LR) was used to find important features.

Results: The study population consisted of 1361 patients. The values of area under the precision-recall curve (AUPRC) were 0.737, 0.710, and 0.562 for the FAST, BE-FAST-1, and BE-FAST-2 models, respectively. The values of AUPRC for the top three ML models were 0.787 for classification and regression tree with undersampling, 0.783 for LR with synthetic minority oversampling technique (SMOTE), and 0.782 for LR with class weighting. Among the ML models, logistic regression and random forest models in general achieved higher values of AUPRC, in particular in those with class weighting or SMOTE to handle class imbalance problem. In addition to the presenting complaint and triage level, age, diastolic blood pressure, body temperature, and pulse rate, were also important features for developing a stroke-alert trigger.

Conclusions: ML techniques significantly improved the performance of prediction models for identification of patients with suspected stroke. Such ML models can be embedded in the electronic triage system for clinical decision support at ED triage.

Abbreviations: AUPRC, area under the precision-recall curve; AUROC, area under the receiver operating characteristic curve; BE-FAST, balance eyes face arm speech time; CART, classification and regression tree; ED, emergency department; EMR, electronic medical record; EVT, endovascular thrombectomy; FAST, Face Arm Speech Test; IVT, intravenous thrombosis; *k*NN, *k*-nearest neighbor; LR, logistic regression; ML, machine learning; RF, random forest; ROC, receiver operating characteristic; SVM, support vector machines; SMOTE, synthetic minority oversampling technique; TIA, transient ischemic attack; TTAS, Taiwan Triage and Acuity Scale.

* Corresponding author.
E-mail address: yhhu@mgt.ncu.edu.tw (Y.-H. Hu).

[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.ijmedinf.2021.104505
Received 25 February 2021; Received in revised form 1 May 2021; Accepted 17 May 2021
Available online 19 May 2021

1. Introduction

Stroke is one of the leading causes of mortality and poses a remarkable disease burden worldwide [1,2]. Approximately 80 % of all strokes are attributed to brain ischemia [3]. Because brain tissue is rapidly lost during brain ischemia, successful treatment of ischemic stroke depends on timely reperfusion of ischemic brain tissue, either by intravenous thrombolysis (IVT) or endovascular thrombectomy (EVT). Current evidence shows that IVT is beneficial for selected patients when treated within 4.5 h of stroke onset [4]. Similarly, earlier treatment with EVT is associated with lower functional disability in large-vessel ischemic stroke, but its benefit became nonsignificant after 7.3 h from onset [5]. Even though careful selection of patients greatly expands the time window of EVT to 24 h [6], patients treated within 12 h of stroke onset may still have better outcomes than those treated beyond 12 h [7].

To expedite the treatment of acute stroke, "code stroke" systems, i.e., organized protocols to facilitate emergency evaluation and treatment of stroke, have been recommended by professional guidelines [8]. Typically, code stroke activities include prompt laboratory examination, neuroimaging, and notification of the on-call neurologist [9–11]. The implementation of code stroke systems effectively improved the quality of care for stroke patients [12,13]. Despite this, code stroke systems may consume lots of medical resources in real-world practice. For example, code strokes were activated an average of five to nine times for each patient who underwent IVT [14,15]. Such disproportionate on-call responsibilities have caused an adverse effect on the recruitment and retention of stroke neurologists [16]. In addition, the potential "burnout" syndrome among stroke neurologists should be addressed properly. Excessive activation of code stroke systems with a low positive predictive value for true stroke may aggravate the situation further. Therefore, an effective triage in the emergency department (ED) to determine whether to activate a code stroke is the key to maintain sustainable operation of code stroke systems.

The Taiwan Triage and Acuity Scale (TTAS), modified from the Canadian Triage and Acuity Scale [17], is a five-level complaint-oriented electronic triage system [18]. During the triage assessment, the triage nurse uses key symptoms, brief medical history, and focused physical examination to determine an appropriate acuity level. By using the Face Arm Speech Test (FAST) [19] to screen for stroke-like symptoms, the acuity level measured by TTAS demonstrated good validity in facilitating timely treatment of acute stroke [20]. However, FAST is unable to identify all stroke patients, in particular those with posterior circulation stroke [21,22]. Fortunately, inclusion of balance (B) and eyes (E) symptoms to FAST, that is, BE-FAST, could increase the sensitivity with a reduction from 14.1% to 4.4% of stroke patients who would have been missed [23].

This study aimed to develop a machine learning (ML) application to identify patients with suspected stroke at ED triage. The application can be embedded in the electronic triage system and used to trigger code stroke. Specifically, this study addressed the following questions: (1) whether BE-FAST performs better in detection of stroke than FAST? (2) Can ML techniques improve the performance further? (3) What are the significant features that help detection of stroke?

2. Materials and methods

2.1. Study population

This was a retrospective study conducted in a 1000-bed teaching hospital. All adult patients suspected as having a stroke or transient ischemic attack (TIA) by ED physicians or triaged with a stroke-related symptom as their presenting complaint in 2016 were included. Patients who presented beyond 12 h of stroke onset or whose electronic medical records (EMRs) were unavailable were excluded.

2.2. Baseline models and triage system

Three rule-based algorithms were treated as the baseline models. The first model used FAST to detect whether patients had a stroke. The second and third ones classified patients as having a stroke when patients presented with one of BE-FAST symptoms as described below.

In TTAS system, each presenting complaint has a set of modifiers, which are used to assign an appropriate acuity level [18]. TTAS is implemented through a computerized decision support tool for Taiwan ED use [18]. A mapping rule was devised to automate the conversion of TTAS presenting complaints to the baseline models (Table 1). The details are given in Supplementary methods in the Supplementary material.

2.3. Dataset

A total of 2469 patients were identified. After excluding 1103 patients presenting beyond 12 h and 5 patients without available EMR data, the remaining 1361 patients comprised the study population (mean age 65.3 years; female 49.1 %). Clinical features at ED triage were collected from EMR, including the presenting complaint, triage level, self-reported medical history (hypertension, diabetes, hyperlipidemia, heart disease, and prior stroke), vital signs, and cardiac rhythm (presence of atrial fibrillation). Table 2 gives the features used for classification. Missing values for nominal and numeric attributes were replaced with the modes and means, respectively.

The class label (outcome) was a diagnosis of stroke or TIA, defined according to the nationwide stroke registry program [24]. For patients who were admitted to the neurology department or assessed by a consultant neurologist at the ED, the final diagnosis was extracted from the discharge summary or the consultation note. For the other patients, the final diagnosis was determined after reviewing clinical symptomatology and neuroimaging findings by a senior stroke neurologist, who was blinded to the clinical features at ED triage. The class label was set to 1 if a patient received a final diagnosis of stroke or TIA, or 0 if otherwise. Among the study population, 433 (31.8 %) patients were determined to have a stroke or TIA.

2.4. Machine learning techniques and data resampling

C4.5, classification and regression tree (CART), *k*-nearest neighbor (*k*NN), random forest (RF), support vector machines (SVM), and logistic regression (LR) were investigated to build prediction models for stroke or TIA. Specifically, J48, SimpleCart, IBK, RandomForest, SMO, and Logistic modules of WEKA 3.8.3 open-source data mining software (Hamilton, New Zealand, www.cs.waikato.ac.nz/ml/weka) were used. The WEKA's implementation of support vector machines transforms nominal attributes into binary ones and normalizes all attributes by default.

Table 1
Baseline models and their mapping from TTAS presenting complaints.

Models	TTAS presenting complaints
FAST	Extremity weakness/symptoms of CVA
BE-FAST-1	Extremity weakness/symptoms of CVA
	Gait disturbance/ataxia
	Visual disturbance, acute or sudden vision change
	Eye pain, acute or sudden vision change
	Diplopia, acute
BE-FAST-2	Extremity weakness/symptoms of CVA
	Gait disturbance/ataxia
	Vertigo, non-positional ± other neurological symptoms
	Visual disturbance, acute or sudden vision change
	Eye pain, acute or sudden vision change
	Diplopia, acute

BE-FAST, Balance, Eyes, Face, Arm, Speech, Time; CVA, cerebrovascular accident; FAST, Face Arm Speech Test; TTAS, Taiwan Triage and Acuity Scale.

Table 2
Clinical features of the study population.

Feature		Data type	Missing
Age, years, mean (SD)	65.3 (15.6)	Numeric	0%
Female, n (%)	668 (49.1)	Nominal	0%
TTAS presenting complaint, n (%)		Nominal	0%
Extremity weakness/symptoms of CVA	432 (31.7)		
Gait disturbance/ataxia	36 (2.6)		
Vertigo, non-positional ± other neurological symptoms	674 (49.5)		
Visual disturbance, acute or sudden vision change	20 (1.5)		
Eye pain, acute or sudden vision change	5 (0.4)		
Diplopia, acute	6 (0.4)		
Other	188 (13.8)		
TTAS triage level, n (%)		Nominal	0%
Level 1	69 (5.1)		
Level 2	1004 (73.8)		
Level 3	285 (20.9)		
Level 4	3 (0.2)		
Hypertension, n (%)	691 (50.8)	Nominal	0%
Diabetes, n (%)	390 (28.7)	Nominal	0%
Hyperlipidemia, n (%)	66 (4.8)	Nominal	0%
Heart disease, n (%)	239 (17.6)	Nominal	0%
Prior stroke, n (%)	247 (18.1)	Nominal	0%
Body temperature, ℃, mean (SD)	36.3 (0.79)	Numeric	0.15 %
Pulse rate, beats per minute, mean (SD)	82.5 (18.3)	Numeric	0.07 %
Respiratory rate, breaths per minute, mean (SD)	19.9 (1.7)	Numeric	0.15 %
Systolic blood pressure, mm Hg, mean (SD)	155.5 (33.0)	Numeric	0%
Diastolic blood pressure, mm Hg, mean (SD)	89.3 (17.6)	Numeric	0%
Atrial fibrillation, n (%)	65 (4.8)	Nominal	0%

CVA, cerebrovascular accident; SD, standard deviation; TTAS, Taiwan Triage and Acuity Scale.

Class imbalance is common in health-related datasets, leading to a bias in the performance of ML algorithms because of their preference towards the majority class. Cost-sensitive learning and data resampling have been widely used to deal with this problem [25,26]. This study explored several resampling methods (Supplementary methods), including undersampling, oversampling, synthetic minority oversampling technique (SMOTE), and class weighting, to test their effect on classifier performance.

2.5. Experiments and statistical analysis

The experimental designs are detailed in the Supplementary methods. In short, two-layer nested cross-validation was used to develop and evaluate classifiers. In the outer layer, the dataset was split into training and holdout test sets in 2:1 ratio. The process of data splitting was repeated 10 times, resulting in 30 training and test set pairs. In the inner layer, 10-fold cross-validation on the training set was performed to find the optimal hyperparameters (Table 3). Finally, by applying the optimal hyperparameters, the whole training set was used to build classifiers, which were tested on the holdout test set.

Because the study aim was to find a model that produce the best precision and recall performance for class label 1, we compared model performance based on the area under the precision-recall curve (AUPRC). Precision (positive predictive value), recall (sensitivity), and the F1 score were calculated. In addition, classification accuracy and the area under the receiver operating characteristic curve (AUROC) were reported for reference.

The average of the estimates from the 30 dataset pairs were compared using the Friedman test for multiple comparison among the models and the Nemenyi test for between-model comparison. Finally, LR was used to evaluate the independent associations between clinical features and the outcome based on the whole dataset. Model discrimination was assessed by AUROC, and model fit was examined by Hosmer-Lemeshow statistics. All statistical analyses were performed using Stata 15.1 (StataCorp, College Station, Texas) and R version 3.6.2 (R Foundation for Statistical Computing, Vienna, Austria). Two-tailed p values <0.05 were considered statistically significant.

Table 3
Machine learning techniques and hyperparameter tuning.

Techniques	Hyperparameters	Range	Increment
C4.5	Confidence factor	0.20 to 0.45	0.05
	Minimum number of instances per leaf	2 to 20	1
CART	Minimum number of instances per leaf	2 to 20	1
*k*NN	Number of neighbors	1 to 20	1
RF	Number of trees	10 to 200	10
SVM	Complexity parameter C	0.01–100.0	×10
	Kernel type	PolyKernel or RBFKernel	NA
LR	Ridge value	1.0E-12 to 10.0	×10

CART, classification and regress tree; *k*NN, *k*-nearest neighbor; LR, logistic regression; NA, not applicable; RF, random forest; SVM, support vector machine.

3. Results

3.1. Model performance

Table 4 lists the performance of baseline models and ML models. FAST achieved an accuracy of 80.1 %, precision of 0.688, recall of 0.686, and F1 score of 0.687, with an AUPRC of 0.737 and AUROC of 0.770. When compared to ML models, FAST ranked in the middle in terms of AUPRC but was inferior to all of them in terms of AUROC (Fig. 1). On the other hand, despite attaining a higher recall (0.711) than FAST, BE-FAST-1 had a lower precision (0.618), lower F1 score (0.661), and lower accuracy (76.8 %) than FAST. BE-FAST-2 had an even higher recall (0.767) at the cost of a much lower precision (0.283), lower F1 score (0.413), and lower accuracy (30.8 %). Its AUPRC and AUROC were the lowest among all models (Fig. 1).

Among ML models, LR and RF generally performed better than the other models in terms of AUPRC (Fig. 2), particularly for those with class weighting or SMOTE to handle class imbalance problem. C4.5 and *k*NN had lower values of AUPRC than the other models. The performance of SVM was not improved by using data resampling except for SMOTE. As for CART, data resampling had various effects on model performance.

The AUPRC values significantly differed across models according to the Friedman test ($p < 0.001$). Among all models, CART with undersampling achieved the highest AUPRC (Fig. 1), followed by LR with SMOTE and LR with class weighting. All the top three models had significantly higher values of AUPRC ($p = 0.001$; $p = 0.001$; $p = 0.003$) and F1 scores (all $p < 0.001$) than FAST. They all showed a higher or comparable recall than FAST, BE-FAST-1, and BE-FAST-2 (all $p < 0.001$ except for $p = 0.365$ between LR with SMOTE and BE-FAST-2). Although the top three models had lower precision than FAST, they all had comparable or higher precision than BE-FAST-1 ($p = 1.000$; $p < 0.001$; $p < 0.001$) and BE-FAST-2 ($p = 0.999$; $p < 0.001$; $p < 0.001$).

3.2. Significant features

The LR analysis on the whole dataset (Table 5) yielded an AUROC of 0.907 (95 % confidence interval 0.890–0.923) with a Hosmer-Lemeshow goodness-of-fit p-value of 0.626, indicating good model fit. Significant features included older age, a complaint of extremity weakness/symptoms of CVA, TTAS triage level 1 or level 2, and higher diastolic blood pressure. Features that were negatively associated with

Table 4
Average performance of various models based on two-layer nested cross-validation.

Models	AUPRC	Precision	Recall	F1 score	AUROC	Accuracy
FAST	0.737	0.688	0.686	0.687	0.770	80.1 %
BE-FAST-1	0.710	0.618	0.711	0.661	0.753	76.8 %
BE-FAST-2	0.562	0.283	0.767	0.413	0.430	30.8 %
C4.5						
Imbalanced	0.701	**0.687**	0.667	0.675	0.862	79.6 %
Undersampling	0.658	0.650	**0.854**	0.737	0.861	80.6 %
Oversampling	0.683	0.654	0.793	0.715	0.859	79.9 %
SMOTE	0.723	0.678	0.785	0.726	**0.873**	**81.2 %**
Class weighting	**0.727**	0.659	0.845	**0.739**	0.867	81.1 %
CART						
Imbalanced	0.712	**0.666**	0.753	0.702	**0.854**	79.9 %
Undersampling	**0.787**	0.627	**0.926**	**0.747**	0.835	80.0 %
Oversampling	0.686	0.636	0.868	0.733	0.836	79.9 %
SMOTE	0.749	0.655	0.869	0.745	0.840	**81.1 %**
Class weighting	0.679	0.639	0.861	0.732	0.840	80.0 %
RF						
Imbalanced	0.766	**0.712**	0.749	0.729	0.891	82.3 %
Undersampling	0.762	0.653	**0.902**	**0.757**	0.892	81.5 %
Oversampling	0.764	0.667	0.873	0.756	0.892	82.0 %
SMOTE	0.770	0.692	0.820	0.750	0.895	**82.6 %**
Class weighting	**0.776**	0.677	0.854	0.755	**0.896**	82.3 %
*k*NN						
Imbalanced	0.690	**0.679**	0.649	0.663	0.861	79.0 %
Undersampling	0.707	0.643	0.882	**0.743**	0.871	**80.6 %**
Oversampling	0.689	0.645	0.858	0.736	0.867	80.4 %
SMOTE	**0.734**	0.645	0.857	0.735	**0.876**	80.4 %
Class weighting	0.715	0.640	**0.885**	0.742	0.871	80.4 %
SVM						
Imbalanced	0.754	**0.685**	0.777	0.725	**0.886**	**81.3 %**
Undersampling	0.711	0.636	**0.924**	0.753	0.873	80.7 %
Oversampling	0.715	0.636	0.923	0.753	0.875	80.7 %
SMOTE	**0.762**	0.648	0.899	0.753	0.883	81.2 %
Class weighting	0.710	0.641	0.921	**0.756**	0.871	81.0 %
LR						
Imbalanced	0.765	**0.699**	0.766	0.730	0.893	**82.0 %**
Undersampling	0.767	0.650	**0.900**	0.754	0.893	81.3 %
Oversampling	0.760	0.650	0.895	0.753	0.892	81.3 %
SMOTE	**0.783**	0.671	0.850	0.750	0.896	**82.0 %**
Class weighting	0.782	0.653	**0.900**	**0.756**	**0.898**	81.5 %

AUPRC, area under the precision-recall curve; AUROC, area under the receiver operating characteristic curve; BE-FAST, Balance, Eyes, Face, Arm, Speech, Time; CART, classification and regress tree; FAST, Face Arm Speech Test; *k*NN, *k*-nearest neighbor; LR, logistic regression; RF, random forest; SMOTE, synthetic minority oversampling technique; SVM, support vector machine.

the outcome included complaints of gait disturbance or ataxia, vertigo (non-positional with or without other neurological symptoms), visual disturbance (acute or sudden vision change), or eye pain (acute or sudden vision change), TTAS triage level 4, higher body temperature, and higher pulse rate.

4. Discussion

The rule-based FAST model performed modestly well according to AUPRC. The rule-based BE-FAST-1 and BE-FAST-2 apparently decreased the missed detection of patients with suspected stroke according to its higher recall than FAST. However, BE-FAST-2 may produce too many false alarms considering its low precision. On the other hand, ML techniques have the potential to improve the identification of patients with suspected stroke according to the F1 scores and values of AUPRC (Table 4). Apart from presenting complaints, several features readily available at ED triage were found useful in the detection of patients with suspected stroke.

4.1. Comparison to previous studies

Despite dissimilar inclusion criteria for study population, the baseline FAST model had a modest performance (recall 0.686; precision 0.688) in identifying patients with stroke or TIA as compared to previous studies (recall 0.79 to 0.97, precision 0.62 to 0.89) [27].

Although a previous study reported BE-FAST had a higher recall (0.957) than FAST (0.859) [23], its precision could not be assessed because that study only included patients with ischemic stroke. BE-FAST-1 and BE-FAST-2 also achieved a higher recall than FAST in the present study but had a lower precision, in particular for BE-FAST-2. The reason may be that BE-FAST-2 considers a presenting complaint of vertigo to be positive evidence of acute stroke. But actually, more than 95 % of patients presenting with a primary symptom of dizziness, vertigo, or imbalance are determined to have a non-cerebrovascular condition [28]. The low precision will inevitably produce excessive false alarms, undermining the utility of these models. Furthermore, strokes presenting with isolated vertigo generally have a low stroke severity [29,30], which obviates the need for IVT or EVT.

4.2. Significant features for developing a stroke-alert trigger

Using the existing complaint-oriented triage system, most of the patients presenting with acute stroke-like symptoms will be assigned to a high TTAS triage level (level 1 or level 2) so that ED physicians can assess these patients within a short time and activate a code stroke in time [20]. This study further identified several significant features that can be used to develop a more sophisticated stroke-alert trigger, including age, diastolic blood pressure, body temperature, and pulse rate.

Age is a major and non-modifiable risk factor for stroke and TIA. The risk of stroke more than doubles every 10 years after the age of 55 years, and up to three quarters of all strokes occur in people aged 65 years or

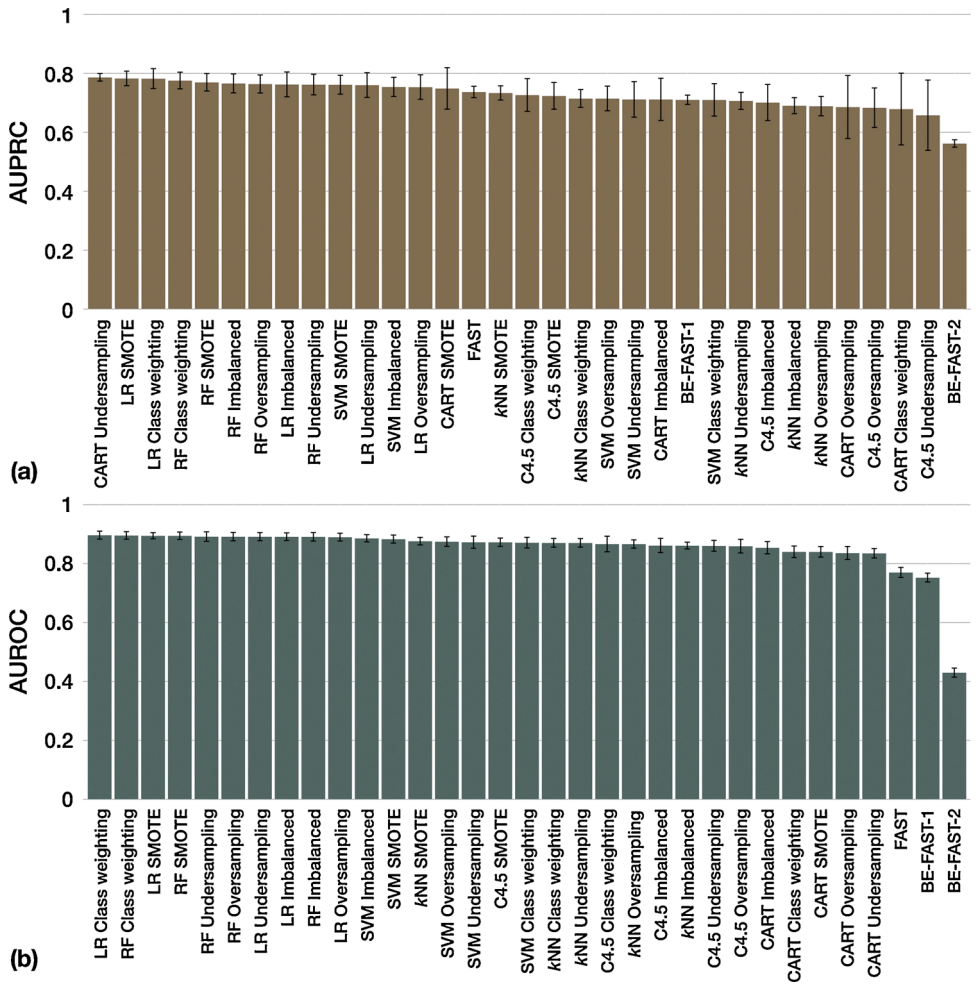


Fig. 1. Performance of various models in predicting a diagnosis of stroke or transient ischemic attack. Prediction models are ordered according to (a) the area under the precision-recall curve (AUPRC) and (b) the area under the receiver operating characteristic curve (AUROC). The error bars indicate one standard deviation.

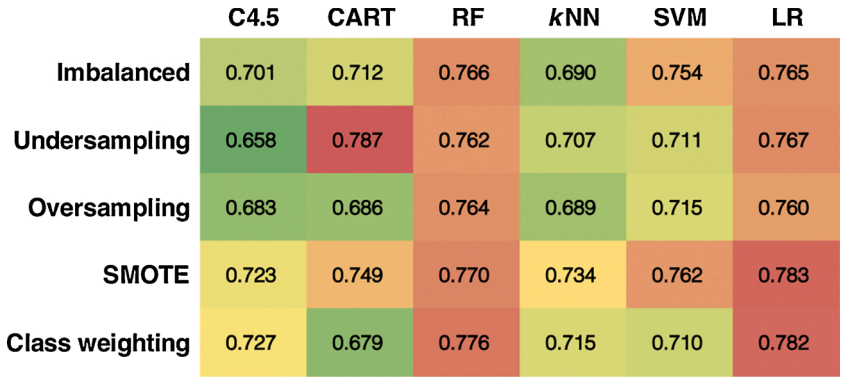


Fig. 2. Heat map showing the area under the precision-recall curve (AUPRC) across machine learning techniques and resampling methods. In the heat map, red color indicates high, yellow intermediate and green low values of AUPRC (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

above [31,32]. Hence, and not surprisingly, older age was found to be a predictor of stroke or TIA in this study. A previous study also reported that age had the highest variable importance value in predicting ischemic stroke in the ED using ML algorithms [33]. On the other hand, high blood pressure is the leading modifiable risk factor for stroke [34]. Moreover, acute hypertensive response with elevation of blood pressure above normal and premorbid values is very common in patients with acute stroke [35,36]. Consequently, elevated diastolic blood pressure could help identify patients with acute stroke.

It is noteworthy that body temperature and pulse rate were negatively associated with a diagnosis of stroke or TIA. Body temperature is usually normal within the first 4 h of stroke even though body temperature may rise and peak at about 36–48 hours after stroke onset [37,38]. High resting pulse rate seems to increase the risk of ischemic stroke [39]. However, there is no evidence indicating whether acute stroke has an effect on pulse rate. At least, the presence of elevated body temperature

Table 5
Logistic regression model based on the whole dataset.

Feature	Coefficient	OR (95 % CI)	*P*
Age, years	0.0154	1.02 (1.00–1.03)	0.009
Female	−0.1754	0.84 (0.60–1.18)	0.310
TTAS presenting complaint			
Extremity weakness/symptoms of CVA	0.7495	2.12 (1.40–3.20)	<0.001
Gait disturbance or ataxia	−0.8741	0.42 (0.18–0.98)	0.046
Vertigo (non-positional ± other neurological symptoms)	−3.8481	0.02 (0.01–0.04)	<0.000
Visual disturbance (acute or sudden vision change)[a]	NA	NA	NA
Eye pain (acute or sudden vision change)[a]	NA	NA	NA
Diplopia, acute	−1.5582	0.21 (0.02–2.43)	0.212
Other	Reference	Reference	Reference
TTAS triage level,			
Level 1	1.2117	3.36 (1.65–6.84)	0.001
Level 2	0.6983	2.01 (1.37–2.94)	<0.001
Level 3	Reference	Reference	Reference
Level 4[a]	NA	NA	NA
Hypertension	0.2496	1.28 (0.90–1.83)	0.166
Diabetes	0.0210	1.02 (0.71–1.47)	0.910
Hyperlipidemia	−0.4477	0.64 (0.32–1.29)	0.213
Heart disease	0.2863	1.33 (0.85–2.09)	0.214
Prior stroke	−0.2537	0.78 (0.52–1.15)	0.204
Body temperature, ℃	−0.2346	0.79 (0.63–0.99)	0.042
Pulse rate, beats per minute	−0.0143	0.99 (0.98–1.00)	0.005
Respiratory rate, breaths per minute	−0.0761	0.93 (0.84–1.02)	0.126
Systolic blood pressure, mm Hg	0.0059	1.01 (1.00–1.01)	0.115
Diastolic blood pressure, mm Hg	0.0244	1.02 (1.01–1.04)	0.001
Atrial fibrillation	0.5841	1.79 (0.87–3.71)	0.115

CI, confidence interval; CVA, cerebrovascular accident; NA, not available; OR, odds ratio; TTAS, Taiwan Triage and Acuity Scale.
[a] The variable predicted negative outcome perfectly and was thus removed from the model.

or increased pulse rate in patients with suspected stroke should alert clinicians to consider an alternative diagnosis, such as sepsis or seizure masquerading as stroke [40].

4.3. Issues regarding imbalanced data

This study dealt with the class imbalance problem using various data resampling methods. Although the performance of the resampling methods varied, SMOTE improved the classification performance across all ML techniques (Fig. 2). A previous empirical study also found the SMOTE-based algorithms as the best performing methodologies [41]. In addition to data resampling, cost-sensitive learning and ensemble methods are other possible solutions to the class imbalance problem. Furthermore, ensemble learning algorithms and data resampling can be combined to form a hybrid method, where data are resampled before training each classifier [41]. Some of the hybrid approaches, such as RUSBoost and UnderBagging, are simple and have good performance [42].

AUPRC was used as the principal evaluation metric because the dataset is imbalanced with 31.8 % positive instances. When evaluating binary classifiers on imbalanced datasets, precision-recall curves were found to be more informative than receiver operating characteristic (ROC) curves [43]. ROC curves could even lead to false conclusions about the reliability of classification performance in the context of imbalanced datasets [43]. A simulation study found that AUROC and AUPRC summarize classification performance equally well in balanced datasets, whereas AUPRC reflected the discriminatory ability of classifiers better than AUROC in imbalanced datasets [44]. Furthermore, precision-recall curves differ from ROC curves in that the number of true negatives is not used for making precision-recall curves [45]. Because this study did not include all ED patients who had no stroke or TIA and were correctly classified as such, AUPRC should be the most reasonable evaluation metric.

4.4. Clinical applications and significance

The application developed in this study differs from existing screening tools in that ML models can estimate the probability of how likely a patient has a stroke. In this way, the cut-off value of probability can be set to reflect the need in different scenarios, which depends on the intended purpose of the application and the trade-off between a false positive prediction and a false negative prediction of stroke. A false negative result, where a patient with true stroke is misclassified as non-stroke and is thus not treated properly, may be considered medical negligence or malpractice, leading to medical disputes and litigation [46,47]. On the other hand, a false positive result, where a patient with a non-stroke condition is misclassified as stroke, may generate futile activation of code stroke and cause burnout of stroke neurologists [14,16].

Under the consideration mentioned above, models with a higher AUPRC are preferred in order to minimize both false positive and false negative results. An electronic triage system equipped with such models can facilitate triage nurses to effectively identify patients likely to have a stroke or TIA, thus enabling timely and appropriate activation of code stroke systems. In EDs where the stroke neurology workforce is adequate, the cut-off value of probability of the model can be set to provide a relatively high recall so that most stroke patients can be identified. In contrast, when stroke neurologists are insufficient in number, the model can be set to have a relatively high precision to avoid overactivation of code stroke systems.

4.5. Limitations

Several limitations are worth noting. First, this is a single-site study and the generalizability of the ML models is unknown and awaits further examination in other settings. Second, the feature importance provided by the LR analysis is relevant only for linear models and may be different from those obtained by decision tree algorithms. Third, prehospital factors, such as mode of transportation to the hospital and diagnosis by emergency medical services, were not considered. Although these factors possibly differ between patients with stroke and those with non-stroke conditions, they are not routinely recorded in the EMR database. Finally, self-reported medical history may be unreliable [48], with variable accuracy across diseases [49]. Future studies may consider collecting medical history directly from the EMR using validated algorithms [50].

5. Conclusions

The widely used FAST had a modest recall and precision in identifying patients with stroke or TIA. BE-FAST achieved a significantly higher recall than FAST but at the cost of a lower precision. Using ML techniques, the performance of prediction models can be markedly improved in terms of AUPRC. In addition to the presenting complaint

and triage level, several significant features, including age, diastolic blood pressure, body temperature, and pulse rate, were found useful for developing a stroke-alert trigger. ML techniques can be used to facilitate ED triage of patients with suspected stroke and may hopefully improve the quality of acute stroke care without causing burnout among stroke neurologists.

Summary table

What was already known on the topic

- While the implementation of code stroke systems can effectively improve the quality of care for stroke patients, an effective triage system in the emergency department (ED) is the key to maintain sustainable operation of code stroke systems.
- The Face Arm Speech Test (FAST) examines facial weakness, arm weakness, and speech disturbance to screen for stroke-like symptoms, but it fails to identify all stroke patients, especially those with posterior circulation stroke.
- The inclusion of balance (B) and eyes (E) symptoms to FAST (i.e., BE-FAST) may increase the sensitivity of identifying stroke patients.
- Research on machine learning (ML) techniques to identify suspected stroke at ED triage has received little attention.

What this study added to our knowledge

- An electronic triage system equipped with ML-based models can facilitate ED triage of patients with suspected stroke, thus enabling timely and appropriate activation of code stroke systems.
- BE-FAST achieved a higher recall than FAST but had a lower precision.
- Classification and regression tree with undersampling technique has significantly higher values of AUPRC and F1 scores than FAST and a higher recall than both FAST and BE-FAST.
- This study identified a number of significant features including age, diastolic blood pressure, body temperature, and pulse rate, which can be used to develop a more sophisticated stroke-alert trigger.

Ethical approval

The study protocol was approved by the Ditmanson Medical Foundation Chia-Yi Christian Hospital Institutional Review Board (CYCH-IRB No.106017).

Informed consent

The requirement for informed consent was waived due to the retrospective design.

Data sharing statement

The complete datasets and codes can be downloaded at the following link: https://osf.io/jkg9t/.

Funding

This research was supported in part by the Ditmanson Medical Foundation Chia-Yi Christian Hospital Research Program (grant number R109-37-3).

Authors' contribution

All authors had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. Study concept and design: All authors. Acquisition of data: SFS and LCH. Analysis and interpretation of data: All authors. Drafting of the manuscript: SFS and LCH. Critical revision of the manuscript for important intellectual content: All authors. Study supervision: YHH.

Declaration of Competing Interest

The authors report no declarations of interest.

Acknowledgements

We would like to thank Ms. Li-Ying Sung for English language editing.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.ijmedinf.2021.104505.

References

[1] GBD 2017 Causes of Death Collaborators, Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980-2017: a systematic analysis for the Global Burden of Disease Study 2017, Lancet 392 (2018) 1736–1788, https://doi.org/10.1016/S0140-6736(18)32203-7.

[2] GBD 2017 DALYs and HALE Collaborators, Global, regional, and national disability-adjusted life-years (DALYs) for 359 diseases and injuries and healthy life expectancy (HALE) for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017, Lancet 392 (2018) 1859–1922, https://doi.org/10.1016/S0140-6736(18)32335-3.

[3] G.C. Fonarow, M.J. Reeves, E.E. Smith, et al., Characteristics, performance measures, and in-hospital outcomes of the first one million stroke and transient ischemic attack admissions in get with the guidelines-stroke, Circ. Cardiovasc. Qual. Outcomes 3 (2010) 291–302, https://doi.org/10.1161/CIRCOUTCOMES.109.921858.

[4] W. Hacke, M. Kaste, E. Bluhmki, et al., Thrombolysis with alteplase 3 to 4.5 hours after acute ischemic stroke, N. Engl. J. Med. 359 (2008) 1317–1329, https://doi.org/10.1056/NEJMoa0804656.

[5] J.L. Saver, M. Goyal, A. van der Lugt, et al., Time to treatment with endovascular thrombectomy and outcomes from ischemic stroke: a meta-analysis, JAMA 316 (2016) 1279–1288, https://doi.org/10.1001/jama.2016.13647.

[6] R.G. Nogueira, A.P. Jadhav, D.C. Haussen, et al., Thrombectomy 6 to 24 hours after stroke with a mismatch between deficit and infarct, N. Engl. J. Med. 378 (2018) 11–21, https://doi.org/10.1056/NEJMoa1706442.

[7] M.J. Alberts, M.D. Ollenschleger, A. Nouh, Dawn of a new era for stroke treatment: implications of the DAWN study for acute stroke care and stroke systems of care, Circulation 137 (2018) 1767–1769, https://doi.org/10.1161/CIRCULATIONAHA.118.033579.

[8] W.J. Powers, A.A. Rabinstein, T. Ackerson, et al., Guidelines for the early management of patients with acute ischemic stroke: 2019 update to the 2018 guidelines for the early management of acute ischemic stroke: a guideline for healthcare professionals from the American Heart Association/American Stroke Association, Stroke 50 (2019) e344–e418, https://doi.org/10.1161/STR.0000000000000211.

[9] S.-F. Sung, C.-T. Ong, C.-S. Wu, et al., Increased use of thrombolytic therapy and shortening of in-hospital delays following acute ischemic stroke: experience on the establishment of a primary stroke center at a community hospital, Acta Neurol. Taiwan. 19 (2010) 246–252.

[10] A.L. Ford, J.A. Williams, M. Spencer, et al., Reducing door-to-needle times using Toyota's lean manufacturing principles and value stream analysis, Stroke 43 (2012) 3395–3398, https://doi.org/10.1161/STROKEAHA.112.670687.

[11] C.-H. Chen, S.-C. Tang, L.-K. Tsai, et al., Stroke code improves intravenous thrombolysis administration in acute ischemic stroke, PLoS One 9 (2014) e104862, https://doi.org/10.1371/journal.pone.0104862.

[12] N. El Husseini, L.B. Goldstein, "Code stroke": hospitalized versus emergency department patients, J. Stroke Cerebrovasc. Dis. 22 (2013) 345–348, https://doi.org/10.1016/j.jstrokecerebrovasdis.2011.09.012.

[13] C.D. Kassardjian, J.D. Willems, K. Skrabka, et al., In-patient code stroke: a quality improvement strategy to overcome knowledge-to-action gaps in response time, Stroke 48 (2017) 2176–2183, https://doi.org/10.1161/STROKEAHA.117.017622.

[14] S.-F. Sung, M.-C. Tseng, Code stroke: a mismatch between number of activation and number of thrombolysis, J. Formos. Med. Assoc. 113 (2014) 442–446, https://doi.org/10.1016/j.jfma.2012.07.029.

[15] M.-C. Sun, P.-J. Hsiao, Time cost of a nonclosing intravenous thrombolysis service for acute ischemic stroke, J. Formos. Med. Assoc. 114 (2015) 910–915, https://doi.org/10.1016/j.jfma.2015.07.005.

[16] E.J. Kenton, A. Culebras, P.B. Fayad, et al., Impact of stroke call on the stroke neurology workforce in the United States: possible challenges and opportunities, J. Stroke Cerebrovasc. Dis. 27 (2018) 2019–2025, https://doi.org/10.1016/j.jstrokecerebrovasdis.2018.02.063.

[17] M.J. Bullard, B. Unger, J. Spence, et al., Revisions to the canadian emergency department triage and acuity scale (CTAS) adult guidelines, CJEM 10 (2008) 136–151, https://doi.org/10.1017/s1481803500009854.
[18] C.-J. Ng, Z.-S. Yen, J.C.H. Tsai, et al., Validation of the Taiwan triage and acuity scale: a new computerised five-level triage system, Emerg. Med. J. 28 (2010) 1026–1031, https://doi.org/10.1136/emj.2010.094185.
[19] J. Harbison, O. Hossain, D. Jenkinson, et al., Diagnostic accuracy of stroke referrals from primary care, emergency room physicians, and ambulance staff using the face arm speech test, Stroke 34 (2003) 71–76.
[20] S.-F. Sung, Y.-C. Huang, C.-T. Ong, et al., Validity of a computerised five-level emergency triage system for patients with acute ischaemic stroke, Emerg. Med. J. 30 (2013) 454–458, https://doi.org/10.1136/emermed-2012-201423.
[21] A.M. Nor, C. McAllister, S.J. Louw, et al., Agreement between ambulance paramedic- and physician-recorded neurological signs with Face Arm Speech Test (FAST) in acute stroke patients, Stroke 35 (2004) 1355–1359, https://doi.org/10.1161/01.STR.0000128529.63156.c5.
[22] G. Gulli, H.S. Markus, The use of FAST and ABCD2 scores in posterior circulation, compared with anterior circulation, stroke and transient ischemic attack, J. Neurol. Neurosurg. Psychiatr. 83 (2012) 228–229, https://doi.org/10.1136/jnnp.2010.222091.
[23] S. Aroor, R. Singh, L.B. Goldstein, BE-FAST (Balance, Eyes, Face, Arm, Speech, Time): reducing the proportion of strokes missed using the FAST mnemonic, Stroke 48 (2017) 479–481, https://doi.org/10.1161/STROKEAHA.116.015169.
[24] F.-I. Hsieh, L.-M. Lien, S.-T. Chen, et al., Get with the Guidelines-Stroke performance indicators: surveillance of stroke care in the Taiwan stroke registry: get with the guidelines-stroke in Taiwan, Circulation 122 (2010) 1116–1123, https://doi.org/10.1161/CIRCULATIONAHA.110.936526.
[25] R.R. Rao, K. Makkithaya, Learning from a class imbalanced public health dataset: a cost-based comparison of classifier performance, Int. J. Electr. Comput. Eng. 7 (2017) 2215–2222, https://doi.org/10.11591/ijece.v7i4.pp2215-2222.
[26] J. Chen, J. Lalor, W. Liu, et al., Detecting hypoglycemia incidents reported in patients' secure messages: using cost-sensitive learning and oversampling to reduce data imbalance, J. Med. Internet Res. 21 (2019) e11990, https://doi.org/10.2196/11990.
[27] M. Rudd, D. Buck, G.A. Ford, et al., A systematic review of stroke recognition instruments in hospital and prehospital settings, Emerg. Med. J. 33 (2016) 818–822, https://doi.org/10.1136/emermed-2015-205197.
[28] B.B. Navi, H. Kamel, M.P. Shah, et al., Application of the ABCD2 score to identify cerebrovascular causes of dizziness in the emergency department, Stroke 43 (2012) 1484–1489, https://doi.org/10.1161/STROKEAHA.111.646414.
[29] R.R. Leker, T.B. Hur, J.M. Gomori, et al., Incidence of DWI-positive stroke in patients with vertigo of unclear etiology, preliminary results, Neurol. Res. 35 (2013) 123–126, https://doi.org/10.1179/1743132812Y.0000000114.
[30] A.M. Elhfnawy, M. Abd El-Raouf, J. Volkmann, et al., Relation of infarction location and volume to vertigo in vertebrobasilar stroke, Brain Behav. 10 (2020) e01564, https://doi.org/10.1002/brb3.1564.
[31] B. Ovbiagele, L.B. Goldstein, R.T. Higashida, et al., Forecasting the future of stroke in the United States: a policy statement from the American Heart Association and American Stroke Association, Stroke 44 (2013) 2361–2375, https://doi.org/10.1161/STR.0b013e31829734f2.
[32] M. Yousufuddin, N. Young, Aging and ischemic stroke, Aging (Albany NY) 11 (2019) 2542–2544, https://doi.org/10.18632/aging.101931.
[33] A.K. Arslan, C. Colak, M.E. Sarihan, Different medical data mining approaches based prediction of ischemic stroke, Comput. Methods Programs Biomed. 130 (2016) 87–92, https://doi.org/10.1016/j.cmpb.2016.03.022.
[34] K.S. Hong, Blood pressure management for stroke prevention and in acute stroke, J. Stroke 19 (2017) 152–165, https://doi.org/10.5853/jos.2017.00164.
[35] A.I. Qureshi, Acute hypertensive response in patients with stroke: pathophysiology and management, Circulation 118 (2008) 176–187, https://doi.org/10.1161/CIRCULATIONAHA.107.723874.
[36] J.P. Appleton, N. Sprigg, P.M. Bath, Blood pressure management in acute stroke, Stroke Vasc. Neurol. 1 (2016) 72–82, https://doi.org/10.1136/svn-2016-000020.
[37] A.A. Wong, J.P. Davis, P.J. Schluter, et al., The time course and determinants of temperature within the first 48 h after ischaemic stroke, Cerebrovasc. Dis. 24 (2007) 104–110, https://doi.org/10.1159/000103124.
[38] B. Karaszewski, P.G.R. Thomas, M.S. Dennis, et al., Temporal profile of body temperature in acute ischemic stroke: relation to stroke severity and outcome, BMC Neurol. 12 (2012) 123, https://doi.org/10.1186/1471-2377-12-123.
[39] W.T. O'Neal, W.T. Qureshi, S.E. Judd, et al., Heart rate and ischemic stroke: the REasons for Geographic and Racial Differences in Stroke (REGARDS) study, Int. J. Stroke 10 (2015) 1229–1235, https://doi.org/10.1111/ijs.12620.
[40] P.J. Hand, J. Kwan, R.I. Lindley, et al., Distinguishing between stroke and mimic at the bedside: the brain attack study, Stroke 37 (2006) 769–775, https://doi.org/10.1161/01.STR.0000204041.13466.4c.
[41] V. López, A. Fernández, S. García, et al., An insight into classification with imbalanced data: empirical results and current trends on using data intrinsic characteristics, Inform. Sci. 250 (2013) 113–141, https://doi.org/10.1016/j.ins.2013.07.007.
[42] M. Galar, A. Fernández, E. Barrenechea, et al., A review on ensembles for the class imbalance problem: bagging-, boosting-, and hybrid-based approaches, IEEE Transactions syst, Man Cybern. Part C Appl. Rev. 42 (2012) 463–484, https://doi.org/10.1109/tsmcc.2011.2161285.
[43] T. Saito, M. Rehmsmeier, The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets, PLoS One 10 (2015) e0118432, https://doi.org/10.1371/journal.pone.0118432.
[44] B. Ozenne, F. Subtil, D. Maucort-Boulch, The precision–recall curve overcame the optimism of the receiver operating characteristic curve in rare diseases, J. Clin. Epidemiol. 68 (2015) 855–859, https://doi.org/10.1016/j.jclinepi.2015.02.010.
[45] J. Davis, M. Goadrich, The relationship between precision-recall and ROC curves, Proceedings of the 23rd international conference on Machine learning, ICML 06 (2006) 233–240, https://doi.org/10.1145/1143844.1143874.
[46] M.I. Weintraub, Thrombolysis (tissue plasminogen activator) in stroke: a medicolegal quagmire, Stroke 37 (2006) 1917–1922, https://doi.org/10.1161/01.STR.0000226651.04862.da.
[47] A. Bhatt, A. Safdar, D. Chaudhari, et al., Medicolegal considerations with intravenous tissue plasminogen activator in stroke: a systematic review, Stroke Res. Treat. 2013 (2013) 562564–562566, https://doi.org/10.1155/2013/562564.
[48] D.M. Tisnado, J.L. Adams, H. Liu, et al., What is the concordance between the medical record and patient self-report as data sources for ambulatory care? Med. Care 44 (2006) 132–140.
[49] S.S. Merkin, K. Cavanaugh, J.C. Longenecker, et al., Agreement of self-reported comorbid conditions with medical and physician reports varied by disease among end-stage renal disease patients, J. Clin. Epidemiol. 60 (2007) 634–642, https://doi.org/10.1016/j.jclinepi.2006.09.003.
[50] K.P. Liao, T. Cai, G.K. Savova, et al., Development of phenotype algorithms using electronic medical records and incorporating natural language processing, BMJ 350 (2015) h1885, https://doi.org/10.1136/bmj.h1885.