



Acute ischaemic stroke prediction from physiological time series patterns

Qing Zhang,¹ Yang Xie,² Pengjie Ye,¹ Chaoyi Pang¹

1. Australian e-Health Research Centre/CSIRO ICT Centre

2. Graduate School of Biomedical Engineering, The University of New South Wales

RESEARCH

Please cite this paper as: Zhang Q, Xie Y, Ye PJ, Pang CY. Acute ischaemic stroke prediction from physiological time series patterns. AMJ 2013, 6, 5, 280-286. <http://doi.org/10.21767/AMJ.2013.1650>

Corresponding Author:

Qing Zhang
Lvl 5, UQ Health Sciences Building 901/16
Royal Brisbane and Women's Hospital
Herston 4029
Queensland, Australia
[Email: Qing.Zhang@csiro.au](mailto:Qing.Zhang@csiro.au)

Abstract

Background

Stroke is one of the major diseases with human mortality. Recent clinical research has indicated that early changes in common physiological variables represent a potential therapeutic target, thus the manipulation of these variables may eventually yield an effective way to optimise stroke recovery.

Aims

We examined correlations between physiological parameters of patients during the first 48 hours after a stroke, and their stroke outcomes after three months. We wanted to discover physiological determinants that could be used to improve health outcomes by supporting the medical decisions that need to be made early on a patient's stroke experience.

Method

We applied regression-based machine learning techniques to build a prediction algorithm that can forecast three-month outcomes from initial physiological time series data during the first 48 hours after stroke. In our method, not only did we use statistical characteristics as traditional prediction features, but we also adopted trend patterns of time series data as new key features.

Results

We tested our prediction method on a real physiological data set of stroke patients. The experiment results revealed

an average high precision rate: 90%. We also tested prediction methods only considering statistical characteristics of physiological data, and concluded an average precision rate: 71%.

Conclusion

We demonstrated that using trend pattern features in prediction methods improved the accuracy of stroke outcome prediction. Therefore, trend patterns of physiological time series data have an important role in the early treatment of patients with acute ischaemic stroke.

Key Words

Stroke outcome prediction, Time series data, Machine Learning

What this study adds:

1. The accuracy of existing stroke prediction methods, based on statistical characteristics of certain physiological variables such as blood pressure, glucose, is unsatisfactory due to vague understandings of effects and function domains of those physiological determinants.
 2. We propose the use of trend patterns of physiological time series data as new key features in predicting three-month stroke outcomes.
 3. Understanding clinical functions behind physiological trend patterns will help decision-making in early treatment of patients with stroke.
-

Background

Stroke is a common cause of human mortality, especially after ischaemic heart disease.¹ The World Health Organisation (WHO) defined stroke as "rapidly developing clinical signs of local (or global) disturbance of cerebral function, with symptoms lasting more than 24 hours or leading to death, and with no apparent cause other than of vascular origin".² Recent research revealed a strong association between physiological homeostasis and outcomes of acute ischaemic stroke. Specifically, the correlations between blood pressure (BP) and stroke outcomes have been widely reported in the literature. Current guidelines have discouraged significant decreases in BP during the first hours after admission. These decreases have been correlated with poor outcomes as measured by



the Canadian Stroke Scale or modified Rankin Score (mRS) at three months.³ Moreover, extreme hypertension and hypotension on admission have also been associated with adverse outcomes in patients presenting with acute stroke.⁴ For example, high baseline systolic BP is inversely associated with favourable outcomes as measured by the mRS at 90 days with OR=1.22 and (95% CI: 1.01, 1.49).⁵ Descriptive abilities for other periodically retrieved properties of BP within 24 hours of ictus (e.g., maximum, mean, variability) have also been investigated. Yong et al.⁶ reported a strong independent association between statistically descriptive properties and outcome at 30 days after ischaemic stroke. For example, variability of systolic BP is inversely associated with favourable outcome with OR=0.57, (95% CI: 0.35, 0.92).

Research has also shown associations between other physiological variables and stroke outcomes. Abnormalities of blood glucose⁷, heart rate variability,⁸ ECG,⁹ and temperature¹⁰⁻¹² may be predictors of three-month stroke outcomes. For instance, heart rate and ECG are associated with stroke outcomes at three-months:

- Heart Rate Variability: Gujjar et al.⁸ reported that heart rate variability was efficient in predicting stroke outcomes. Specifically, they studied a continuous echocardiogram for 25 patients with acute stroke and concluded that the eye-opening score of Glasgow Coma Scale (GCS) and low-frequency spectral power were factors that were independently predictive of mortality.
- ECG: Christensena et al. reported the relationship between ECG abnormalities and stroke outcomes.⁹ They analysed a large cohort with 692 patients. And they concluded that ECG abnormalities were frequent in acute stroke and may conclude with three-month mortalities.

Thus, understanding determinants of physiological variables may yield an effective and potentially widely applicable range of therapies for optimising stroke recovery, such as abbreviating the duration of ischaemia, preventing deterioration due to post-stroke complications, or preventing further stroke.

Related work

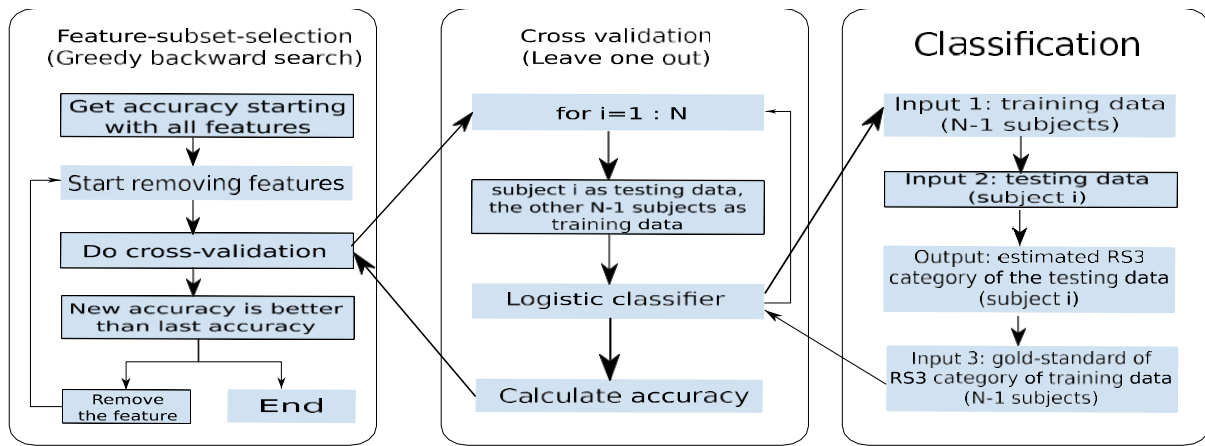
Most of the above analyses, if not all, are based on statistical properties of periodical snapshots of physiological parameters, hourly or daily, up to three months. Whether continuous patterns of physiological stream data, such as data trends, having a similar predictive role remained unknown. Although it is clear that elevated blood pressure levels within 24 hours after stroke has predicted a poor outcome, few studies have investigated the predictive ability of more sophisticated trends (e.g., combined trends of several physiological parameters). Yet this could be an effective way to readily obtain important prognostic information for acute ischaemic stroke patients.

Dawson et al. conducted pioneering work on associating shorter length (approximately 10 minutes) beat-to-beat BP with acute ischaemic stroke outcomes.¹³ They concluded that a poor outcome, assessed by mRS, at 30 days after ischaemic stroke was dependent on stroke subtypes, beat-to-beat diastolic BP, and Mean Arterial Pressure and variability. However, in their study, they still used the average values of continuous BP recordings instead of time series data patterns. A further study on BP investigated detrimental effects of BP reduction in the first 24 hours of acute stroke onset.¹⁴ BP reduction can worsen an already compromised perfusion in the brain tissue. Thus, the avoidance of lowering BP in the early stage after the stroke onset has been suggested. However, further discussion on the relation of higher BP and the outcomes are lacking. Ritter et al. formulated the BP variation by counting threshold violations.⁵ Significant difference in the frequency of upper threshold violation occurrences was observed between different time points after stroke.⁵ Wong (2008) observed temporal patterns from the changing process of physiological variables and also attempted to employ such temporal patterns to explain and predict early stroke outcomes.⁷ These studies motivated our research on mining physiological data patterns as effective predictors of acute ischaemic stroke outcome.



A1

Figure 1: The flowchart of our stroke outcome prediction method



From a technical point of view, mining physiological data patterns can be easily aligned with the time series data classification that is a traditional topic and has attracted intensive study. Although many sophisticated time series data mining techniques exist, we find that most of them are not applicable to our scenarios, mainly due to the incomplete, non-isometric physiological data collected from stroke patients. Therefore, in this paper, we incorporated a simple yet powerful time series data pattern analysing method, namely trend analyses, into our prediction method. By utilising trend features of time series data, together with traditional physiological variables, we designed an efficient algorithm to predict three-month stroke outcomes with high accuracy, as further explored in the following section.

Method

Our prediction method adopted statistical values of physiological parameters and also incorporated the descriptive ability of the physiological patterns as features to predict three-month stroke outcomes. Particularly, we use trend patterns of time series data as new add-on features to form an initial feature set. Then, we applied the logistic regression method to classify stroke patient outcomes into two groups: good versus bad. Cross validation was adopted to obtain an unbiased assessment of classifier performance. Finally, we selected a final feature subset that most accurately predicted three-month stroke outcomes. Figure 1 presents logic flows of our method. Note that we used the Rankin Scale to represent various outcomes at three months after stroke (RS3).¹⁵ Also, there exist three different clinical criteria in defining good/bad outcomes. We ran our prediction algorithm on all criteria and reported empirical study results in the next section.

Construct initial feature set

Generally, five physiological parameters are considered as influential factors on stroke patient outcomes, namely Blood Sugar Level, Diastolic Blood Pressure, Systolic Blood Pressure, Heart Rate and Body Temperature.^{8,9,14} Existing stroke outcome prediction methods assumed a certain parameter as the main feature in their approaches.

However, in our approach, we assumed all five parameters as features in the initial feature set.

Moreover, for each physiological time series data, we computed trend patterns through partitioning into several non-overlapping, continuous blocks. Although many trend and shape detection methods have been reported in the literature,¹⁶ in our application, we only considered a bi-partition on first 48-hour time series data records after stroke. The reasons are:

1. Most available physiological data records are only measured within 48 hours after stroke.
2. Clinical observation and our initial experiments suggested that by setting the granularity level at only two partitions in 48 hours, trend patterns can well represent changes of physiological time series data.

In each partition, accordingly we generated six new features, listed below, to represent the trend pattern:

1. $yChange$: the difference between the value at the end of a trend and the value at the start of a trend
2. $absYChange$: the absolute value of the $yChange$
3. $slope$: the slope of the trend
4. $sign$: the direction of the trend
5. $NumofMeasure$: the number of values in a partition
6. $FreqofMeasure$: the average time interval between measurements, i.e.

$$FreqofMeasure = \frac{TrendLength}{NumofMeasure}$$

Now, the initial feature set comprised values of physiological data and their trend patterns. We next applied the logistical regression method to classify the good/bad stroke outcomes based on this initial feature set.



Logistic Regression Classifier

In statistics, logistic regression is a type of regression analysis used to predict the outcome of a binary dependent variable, a variable that can take only two possible outcomes. Like other forms of regression analyses, logistic regression uses one or more predictor variables, either continuous or categorical. The difference is in the categorisation of the outcome as binary. As in our study, based on our initial feature set, we used logistic regression to predict the outcomes of stroke (*good vs. bad*).

To obtain an unbiased assessment of the classifiers' performance, the Leave-One-Out Cross Validation technique was adopted. Suppose N folds are employed, this technique withholds a subject from the training set for each run to be used in a later test. Once a record has been withheld for testing, the classifier is trained using the remaining N-1 subjects. The withheld subject is then reintroduced for next round of classifications.

Final feature set selection

We used subset selection to find the 'best' feature subset that achieved the highest prediction accuracy. Rather than search through all possible subsets, we used two greedy search strategies: *backward search* and *forward search*:

Backward search: This method started with a set that included every feature and sequentially removed a feature that improved prediction accuracy the most (or decreased it the least) from the current set of features. This process continued until all features had been removed. The intermediate feature subset with the highest performance, compared to all other subsets evaluated, was selected as the final feature set.¹⁷

Forward search: This method attempted to find the optimal subset of features from the pool of available candidate features. Starting with a set that included only one feature from all available features, it sequentially added a new feature that improved the prediction accuracy to a maximum extent in each step. After the selection, removal of a feature from the selected features was applied. The process of possible feature addition, followed by possible feature removal, was iterated until the selected feature set converges.¹⁸

Results

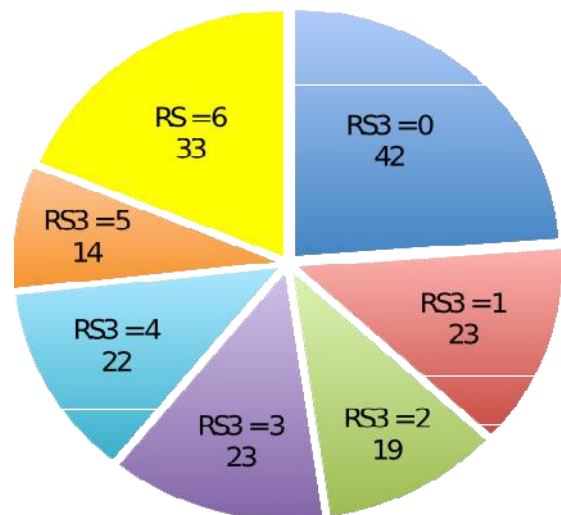
We reported results of testing our prediction method on a real data set of stroke patients. First, we introduced the physiological datasets of stroke patients and the good/bad criteria used in our study. Then we reported prediction accuracy comparisons under different criteria. Our study was

approved by an ethics committee from the related institution.

Experimental data sets

A cohort with 157 patients with acute ischaemic stroke was recruited. Patients who presented to the Emergency Department of the Royal Brisbane and Women's Hospital, an Australian tertiary referral teaching hospital, within 48 hours of stroke or existing inpatients with an intercurrent stroke were enrolled prospectively. Important physiological parameters, such as blood pressure, were recorded at least every four hours from the time of admission until 48 hours after the stroke. These values were used as outcome variables in the analyses. Measurements from patients who died during these first 48 hours were also included in the analyses. Furthermore, some demographic and related data were also collected such as age and gender. The age range of these 157 patients was from 16- to 92 years of age with a median age of 75 years old. The patient distribution based on different values of RS3 is presented in Figure 2.

Figure 2: Patient distributions on different values of RS3



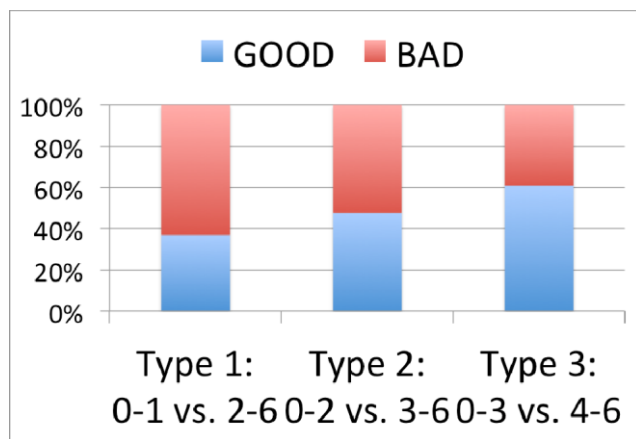
Ranking Scales	
0-	No symptoms
1	No significant disability
2	Slight disability
3	Moderate disability
4	Moderately severe disability
5	Severe disability
6	Dead



Classification criteria

As shown in Figure 2, RS3 score varied from 0 to 6 points. RS3=6 indicated that the subject had died by three months; an RS3=0 indicated that the subject recovered quite well by three months. Based on RS3 values, patient outcomes can be divided into good/bad groups with different classification criteria. Figure 3 illustrates patient distributions under three types of grouping criteria. That is Type 1 assuming RS3 between 0 and 1 as good outcomes. Type 2 assumes RS3 between 0 and 2 as good outcomes. And Type 3 assumes RS3 between 0 and 3 as good outcomes. That is Type 1 assuming RS3 between 0 and 1 as good outcomes. Type 2 assumes RS3 between 0 and 2 as good outcomes. And Type 3 assumes RS3 between 0 and 3 as good outcomes.

Figure 3: Distributions of patients with good/bad outcomes under various criteria (Y-axis: distribution percentage)



Prediction accuracy comparisons

With the prediction techniques previously described, we ran analyses on all above three types of grouping criteria to test our stroke outcome prediction algorithms. We noticed that 'backward search' generated more accurate prediction results, which will thus be used as our default feature set search strategy. Figure 4 shows prediction accuracy comparisons under all three types of grouping criteria. In Figure 5, we also evaluated the benefits of adding trend patterns as new prediction features. We noticed that compared against predictions with only temporal and statistical features, on average these trend pattern feature based predictions can achieve 20 per cent increase of estimation accuracy.

Discussion

The inclusion of trend patterns as prediction features in our algorithm achieved a higher precision rate as well as a good recall rate.

Compared against traditional prediction methods that did

not consider trend patterns of physiological parameters, we demonstrated that trend patterns play an important role in the improvement of prediction accuracy.

However, our cohorts were relatively small. That is also why we only try our methods on dichotomous classification in this study. We anticipate clinical trials on larger cohorts to validate our prediction tool, to test prediction accuracy especially on RS3-score based classification.

Figure 4: Prediction accuracy comparisons under three types of grouping criteria

Precision: 86% Recall: 88%

		Estimated	
		Good	Bad
True	Good	56	9
	Bad	8	100

Type 1

Precision: 94% Recall: 90%

		Estimated	
		Good	Bad
True	Good	76	5
	Bad	9	83

Type 2

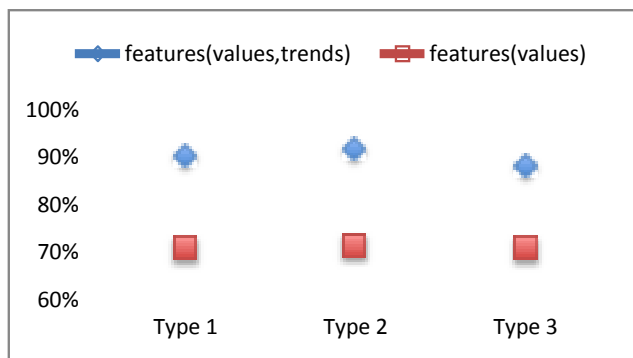
Precision: 93% Recall: 88%

		Estimated	
		Good	Bad
True	Good	97	7
	Bad	13	56

Type 3

We also anticipate new collaborations with healthcare professionals to determine the clinical truth behind those significant physiological trend patterns used in our prediction methods. We believe this will greatly benefit clinical treatments for acute ischaemic stroke.

Figure 5: Including trend patterns as prediction features improves prediction accuracy (Y-axis: prediction accuracy)



Conclusion

In this paper, we have described novel algorithms to predict 3-month stroke outcomes from records of physiological parameters during the 48 hours after stroke. We have quantified improvements through the inclusion of physiological trend patterns as features in our algorithms. We believe that these trends play an important role on early clinical treatments of stroke patients. The efficiency and accuracy of our algorithm have also been demonstrated through experiments on a real data set of stroke patients.

References

1. AIHW. Australia's health 2006. Australia's health no. 10. Cat. no. AUS 73. Canberra: AIHW, 2006.
2. WHO MONICA Project Principal Investigators. The World Health Organization MONICA Project (monitoring trends and determinants in cardiovascular disease): a major international collaboration. WHO MONICA Project Principal Investigators. *J Clin Epidemiol.* 1988;41(2):105-14.
3. Castillo J, Leira R, Garcia MM, Serena J, Blanco M, Davalos A. Blood pressure decrease during the acute phase of ischemic stroke is associated with brain injury and poor stroke outcome. *Stroke.* 2004 Feb;35(2):520-6.
4. Ahmed N, Nasman P, Wahlgren NG. Effect of intravenous nimodipine on blood pressure and outcome after acute stroke. *Stroke.* 2000 Jun;31(6):1250-5.
5. Ritter MA, Kimmeyer P, Heuschmann PU, Dziewas R, Dittrich R, Nabavi DG, Ringelstein EB. Blood pressure threshold violations in the first 24 hours after admission for acute stroke: frequency, timing, predictors, and impact on clinical outcome. *Stroke.* 2009 Feb;40(2):462-8.
6. Yong M, Kaste M. Association of characteristics of blood pressure profiles and stroke outcomes in the ECASS-II trial. *Stroke.* 2008 Feb;39(2):366-72.
7. Wong A. The Natural History and Determinants of Changes in Physiological Variables after Ischaemic Stroke. PhD Thesis, Brisbane: The University of Queensland. 2009.
8. Gujjar AR, Sathyaprabha TN, Nagaraja D, Thennarasu K, Pradhan N. Heart rate variability and outcome in acute

severe stroke: role of power spectral analysis. *Neurocrit Care.* 2004;1(3):347-53.

9. Christensen H, Fogh Christensen A, Boysen G. Abnormalities on ECG and telemetry predict stroke outcome at 3 months. *J Neurol Sci.* 2005 Jul 15;234(1-2):99-103.
10. Boysen G, Christensen H, Stroke severity determines body temperature in acute stroke. *Stroke.* 2001. 32 (2): p. 413-7.
11. Reaven NL, Lovett JE, Funk SE, Brain injury and fever: hospital length of stay and cost outcomes. *J Intensive Care Med.* 2009. 24(2): p. 131-9.
12. Hajat C, Hajat S, Sharma P, Effects of poststroke pyrexia on stroke outcome : a meta-analysis of studies in patients. *Stroke.* 2000. 31(2): p. 410-4.
13. Dawson SL, Manktelow BN, Robinson TG, Panerai RB, Potter JF. Which parameters of beat-to-beat blood pressure and variability best predict early outcome after acute ischemic stroke? *Stroke.* 2000 Feb;31(2):463-8.
14. Oliveira-Filho J, Silva SC, Trabuco CC, Pedreira BB, Sousa EU, Bacellar A. Detrimental effect of blood pressure reduction in the first 24 hours of acute stroke onset. *Neurology.* 2003 Oct 28;61(8):1047-51.
15. Rankin J. Cerebral vascular accidents in patients over the age of 60. II. Prognosis. *Scott Med J.* 1957 May;2(5):200-15.
16. Ye L, Keogh E. Time series shapelets: a new primitive for data mining. *Proceedings of the 15th ACM SIGKDD.* 2009; p. 947-56.
17. Redmond SJ, Xie Y, Chang D, Basilakis J, Lovell NH. Electrocardiogram signal quality measures for unsupervised telehealth environments. *Physiol Meas.* 2012 Sep;33(9):1517-33.
18. Narayanan MR, Redmond SJ, Scalzi ME, Lord SR, Celler BG, Lovell Ast NH. Longitudinal falls-risk estimation using triaxial accelerometry. *IEEE Trans Biomed Eng.* 2010 Mar;57(3):534-41.

ACKNOWLEDGEMENTS

We would like to thank Dr. Andrew Wong for allowing us accessing the datasets of stroke patients and for providing clinical interpretations and discussions on our prediction results.

We would also like to thank our reviewer's critical insight and hard work on helping us preparing the final version of this paper.

PEER REVIEW

Not commissioned. Externally peer reviewed

CONFLICTS OF INTEREST

The authors declare that they have no competing interests.



ETHICS COMMITTEE APPROVAL

The physiological time series data of stroke patients was obtained with ethics approval from Royal Brisbane and Women's Hospital Metro North Health Service District – Proposal #HREC/10/QRBW/487.