

Acute Ischemic Stroke Prediction from Physiological Time Series Patterns

Qing Zhang^{1,2}, Yang Xie², Pengjie Ye^{1,2}, and Chaoyi Pang²

¹ Australian e-Health Research Centre/CSIRO ICT Centre

² The University of New South Wales

{qing.zhang, pengjie.ye, chaoyi.pang}@csiro.au

yang.xie@unsw.edu.au

Abstract. Stroke is one of the major diseases that can cause human deaths. However, despite the frequency and importance of stroke, there are only a limited number of evidence-based acute treatment options currently available. 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. Nevertheless the accuracy of prediction methods based on statistical characteristics of certain physiological variables, such as blood pressure, glucose, is still far from satisfactory due to vague understandings of effects and function domain of those physiological determinants. Therefore, developing a relatively accurate prediction method of stroke outcome based on justifiable determinants becomes more and more important to the decision of the medical treatment at the very beginning of the stroke. In this work, we utilize machine learning techniques to find correlations between physiological parameters of stroke patient during 48 hours after stroke, and their stroke outcomes after three months. Our prediction method not only incorporates statistical characteristics of physiological parameters, but also considers physiological time series patterns as key features. Experiment results on real stroke patients' data indicate that our method can greatly improve prediction accuracy to a high precision rate of 94%, as well as a high recall rate of 90%.

Keywords: Stroke, Outcome Prediction, Time Series Data, Machine Learning

1 Introduction

Stroke is a common cause of human death and is a major cause of death after ischemic heart disease [1]. The World Health Organisation (WHO) defines it 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 years research reveals a strong association between physiological homeostasis and outcomes of Acute Ischemic Stroke. Thus understanding determinants of physiological variables, such

Acute Ischemic Stroke Prediction

as blood pressure, temperature and blood glucose levels, may eventually yield an effective and potentially widely applicable range of therapies for optimising stroke recovery, such as abbreviating the duration of ischaemia, preventing further stroke, or preventing deterioration due to post-stroke complications.

The correlations between blood pressure and stroke outcomes have been widely studied in the literature. It is stated in current guidelines that a significant decrease of BP during the first hours after admission should be avoided, as it correlates with poor outcomes, measured by Canadian Stroke Scale or modified Rankin Score (mRS), at 3 months [10]. Extreme hypertension and hypotension on admission have also been associated with adverse outcome in acute stroke patients [11]. BP values, periodically monitored within the first 72 hours after admission, demonstrate that extreme values still correlate with unfavoured outcomes [9]. For example, high baseline of systolic BP is inversely associated with favourable outcome assessed on mRS at 90 days with OR=1.220 and (95% CI: 1.01 to 1.49). Other periodically retrieved statistical properties of BP within 24 hours of ictus, such as maximum, mean, variability etc., have also been investigated. Yong et al. [12] report strong independent association between those properties and the outcome at 30 days after ischemic stroke. For example, variability of systolic BP is inversely associated with favourable outcome with OR=0.57, (95% CI: 0.35 to 0.92).

Research also shows associations between other physiological variables and stroke outcomes. Abnormalities of blood glucose, heart rate variability, ECG and temperature may be predictors of 3-month stroke outcome.

Most of the above analyses are based on periodically recorded physiological parameters, hourly or daily, up to 3 months. Whether continuous data patterns, such as data trends, have a similar predictive role is still uncertain. Although it is clear that the after stroke elevated 24-hours blood pressure levels predict 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 ischemic stroke patients. Dawson et al did pioneering works on associating shorter length (around 10 minutes) beat-to-beat BP with acute ischemic stroke outcomes [8]. They conclude that a poor outcome, assessed by mRS, at 30 days after ischemic stroke is dependent on stroke subtype, beat-to-beat diastolic BP and Mean Arterial Pressure and variability. However in their study, they still use the average values of continuous recordings, instead of time series patterns as predictors. This motivates our research on mining physiological data patterns as effective predictors of acute ischemic stroke outcome.

Obviously mining physiological data patterns can be easily aligned with time series data classification, which is a traditional topic and has attracted intensive studies. Although there exist many sophisticated time series data mining techniques, we find that most of them, if not all, are not applicable to our application scenario, due to the always incomplete, non-isometric physiological data collected from patients. Therefore, in this paper, we incorporate a simple yet powerful time series data pattern analysing method, trend analyses, into

Acute Ischemic Stroke Prediction

our prediction method. By utilising those trend features, together with values of traditional physiological variables, we design an efficient algorithm that can predict 3-month stroke outcome with high accuracy.

In summary, we list our contributions in this paper:

- We propose using trend patterns of physiological time series data as a new set of stroke outcome prediction features,
- We design a novel prediction algorithm which can accurately predict 3-months stroke outcomes with high precision and recall rate, when tested against a real data set.

The rest of this paper is organised as follows. Section 2 introduces works related to stroke outcome predictions. Section 3 presents our prediction methods. Section 4 reports empirical study results. And section 5 concludes this paper with possible future studies.

2 Related Work

The relationship between beat-to-beat blood pressure (BP) and the early outcome after acute ischemic stroke was firstly described in [8].

A further investigation on BP was done in [6], which investigated detrimental effects of blood pressure reduction in the first 24 hours of acute stroke onset. BP reduction is regarded to have the possibility to worsen an already compromised perfusion in the brain tissue and thus not lowering BP in the early stage after the stroke onset is suggested. However, it lacks further discussion on the relation of higher BP and outcome. Ritter et al. formulated the blood pressure variation by counting threshold violations. Significant difference in the frequency of upper threshold violation occurrences was observed between different time points after stroke [9]. Wong observed some temporal patterns from the changing process of some physiological variables and also attempted to employ such temporal patterns to explain and predict the early outcomes [5]. However, due to the limit of candidate feature set considered in those studies, achieving an accurate prediction is fairly unlikely in those scenarios.

Relationships between other physiological variables and stroke outcome have also been studied in literature. Abnormalities of serum osmolarity, temperature, blood glucose, SPO₂ may be predictors of stroke outcomes. More specifically, heart rate and ECG, can be correlated to stroke outcomes at 3-months:

- Heart Rate Variability: Gujjar et al. reported that heart rate variability is efficient in predicting stroke outcome. Specifically they studied continuous echocardiogram of 25 patients with acute stroke and concluded that the eye-opening score of Glasgow Coma Scale and low-frequency spectral power were factors that were independently predictive of mortality [16].
- ECG: The relationship between ECG abnormalities and stroke outcomes were reported by Christensen et al. They analysed a large cohort of 692 patients and predict that ECG abnormalities are frequent in acute stroke and may conclude 3-month mortality [17].

3 Stroke outcomes prediction

Our prediction method adopts statistical values of physiological parameters and also incorporates the descriptive ability of the physiological patterns as features to predict 3-months stroke outcomes. Particularly, we use the trend pattern of time series data as new add-on features to form an initial feature set. Then we apply the logistic regression method to classify stroke patient outcomes into two groups: good vs. bad. Note that there exist different clinical criteria in defining good/bad outcomes. We will report empirical study results on all criteria in the next section. Cross validation is also adopted to obtain an unbiased assessment of classifier performance, by which the physiological determinants can be accurately identified in the last stage. Finally, we select a subset of features that can most accurately predict 3-months stroke outcomes. Figure 1 presents logic flows of our method. We use Rankin Scale to represent various outcomes at 3 months after stroke (RS3) [18].

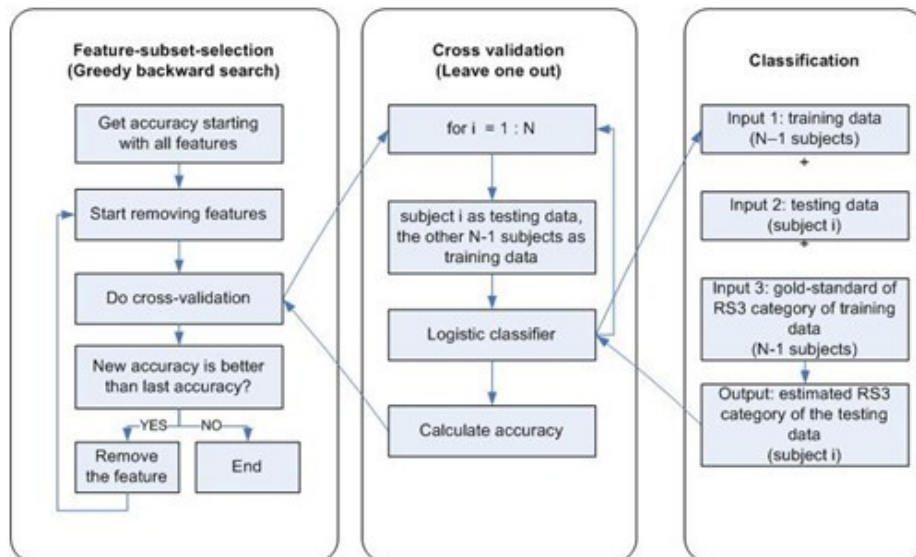


Fig. 1. Stroke outcomes prediction method

3.1 Construct initial feature set

Five physiological parameters are usually considered as influential factors on stroke patient outcomes, namely Blood Sugar Level, Diastolic Blood Pressure, Systolic Blood Pressure, Heart Rate and Body Temperature [6, 16, 17]. Existing stroke outcome predictions always assume a certain parameter as the main

Acute Ischemic Stroke Prediction

feature in their approaches. However in our approach, we will assume all five parameters in the initial feature set.

Moreover, for each physiological parameter, we compute trends through partitioning the time series data into non-overlapping, continuous blocks. Although there exists many trend and shape detection methods in the literature, such as [3], in our application, we simply consider a bi-partition on the first 48-hours time series data records after stroke. The reasons are:

1. most available physiological data records are only within 48-hours after stroke.
2. clinical observation and our initial experiments both suggest that setting the granularity level at having only two partitions in the 48-hours, well represents the physiological time series pattern changes.

In each partition, accordingly we generate 6 new features, as shown 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

$$yChange = y(end_of_trend) - y(start_of_trend)$$

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.

$$FreqofMeasure = \frac{Trend.Length}{NumofMeasure}$$

The initial feature set comprised physiological values and their trend patterns. We apply the logistical regression method to classify the good/bad stroke outcomes based on this initial feature set.

3.2 Logistic Regression Classifier

In statistics, logistic regression is a type of regression analysis used for predicting the outcome of a binary dependent variable (a variable which can take only two possible outcomes, e.g. “yes” vs. “no” or “success” vs. “failure”) based on one or more predictor variables. Like other forms of regression analysis, logistic regression makes use of one or more predictor variables that may be either continuous or categorical. Unlike ordinary linear regression, however, logistic regression is used for predicting binary outcomes rather than continuous outcomes. Logistic regression adopted here is a type of regression analysis used for predicting the outcome of stroke (“good” vs. “bad”) based on features in our initial feature set.

To obtain an unbiased assessment of classifier performance, the Leave-One-Out Cross validation technique is adopted. Suppose N folds are employed, this

Acute Ischemic Stroke Prediction

technique withholds a subject from the training set for each run to later test with. 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 classification.

3.3 Final feature set selection

We use two greedy search strategies to find the best feature subset that can achieve highest prediction accuracy. Specifically, we use backward search and forward search:

backward search : A greedy backward search is performed to identify a near optimum subset of features. Starting with all features, in sequence, the feature which improves prediction accuracy the most (or decreases it the least) is removed from the current set of features and retained as an intermediate feature subset. This is repeated until all features have been removed. The intermediate feature subset which provides the maximum performance, compared to all other subset evaluated, is selected as the final feature set.

forward Search A sequential forward floating search algorithm is used for feature selection, in an attempt to discover the optimal subset of features from the pool of available candidate features. This strategy begins with a forward-selection process, selecting a single feature from the pool of available features, which improves the prediction accuracy most. After this selection, removal of a feature from the set of selected features is considered. The process of possible feature addition, followed by possible feature removal, is iterated until the selected feature set converges.

4 Empirical Study

In this section, we report experiment results through testing our prediction method on a real data set of stroke patients. Firstly, we introduce the physiological data sets of stroke patients and the good/bad criteria used in our study. Then we report prediction accuracy based on various combination of feature sets. Our study was approved by a ethics committee of the related institution.

4.1 Experimental data sets

A cohort of 157 patients with acute ischaemic stroke were recruited. Patients presenting 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 4 hours from the time of admission until 48 hours after the stroke.

Acute Ischemic Stroke Prediction

These values were used as the outcome variable in the analyses. The measurements from patients who died during these first 48 hours were also included in the analyses. Furthermore, some demographic and other stroke-related data were also collected such as the age and gender. The age range of these 157 patients was 16 to 92 years with median age 75 years. The patient distribution based on different values of RS3 is showed in Figure 2.

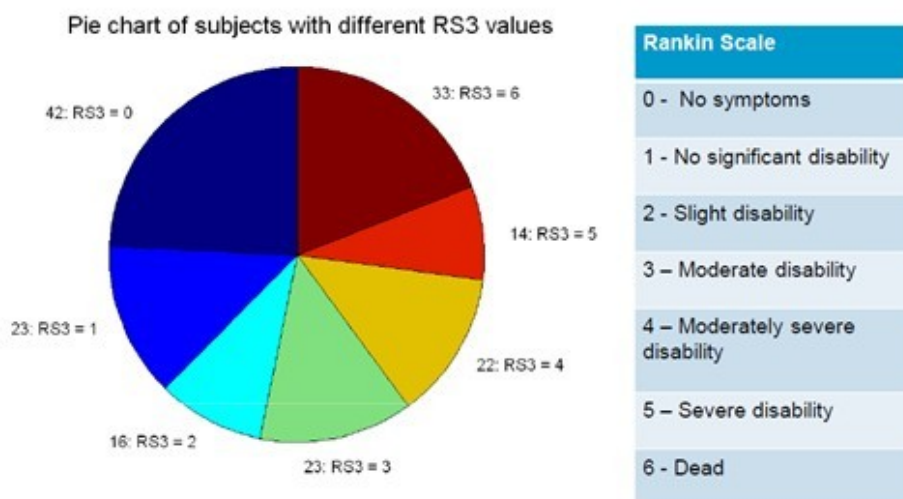


Fig. 2. Patient distributions on values of RS3

4.2 Classification criteria

As shown in Figure 2, RS3 score varies between 0 and 6. Patients with RS3 = 6 means the subject is dead after three months and RS3 = 0 means the subject recovers quite well after three months. Based on RS3 values, patient outcomes can be divided into good/bad groups basing on different grouping criteria. Figure 3 illustrates patient distributions under three type grouping criteria.

4.3 Prediction accuracy comparisons

Applying techniques described in Section 3, we run experiments on various grouping criteria to test our stroke outcome prediction algorithm. We always notice that 'backward search' generates 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 evaluate the efficiency of including trend pattern as prediction

Acute Ischemic Stroke Prediction

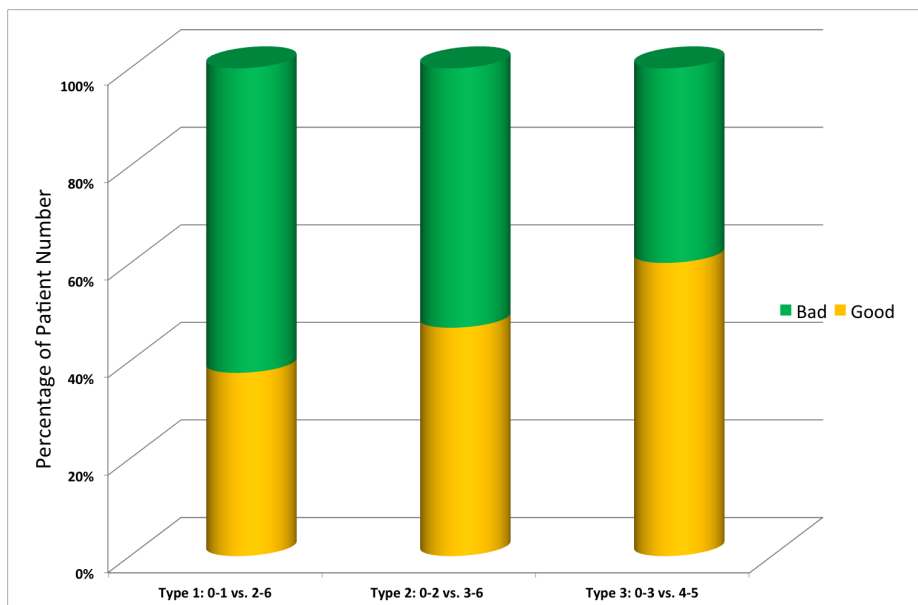


Fig. 3. Good vs Bad outcomes under various criteria

		Estimated	
		Good	Bad
True	Good	56	9
	Bad	8	100

Precision: 86%
Recall: 88%

Type 1

		Estimated	
		Good	Bad
True	Good	76	5
	Bad	9	83

Precision: 94%
Recall: 90%

Type 2

		Estimated	
		Good	Bad
True	Good	97	7
	Bad	13	56

Precision: 93%
Recall: 88%

Type 3

Fig. 4. Prediction Accuracy on different grouping criteria

features. Experiment shows that by adding those simple trend features, the prediction accuracy on all three grouping types is unanimously boosted from 71% to 89~91%.

5 Conclusion

In this paper, we describe novel algorithms to predict three months stroke outcomes. We have quantified the great improvements brought by including physiological data trend patterns as features of a classifier. We believe that these trends play important roles on three months outcomes of stroke patients. The efficiency and accuracy of our algorithm have also been demonstrated through our experiments.

Acute Ischemic Stroke Prediction

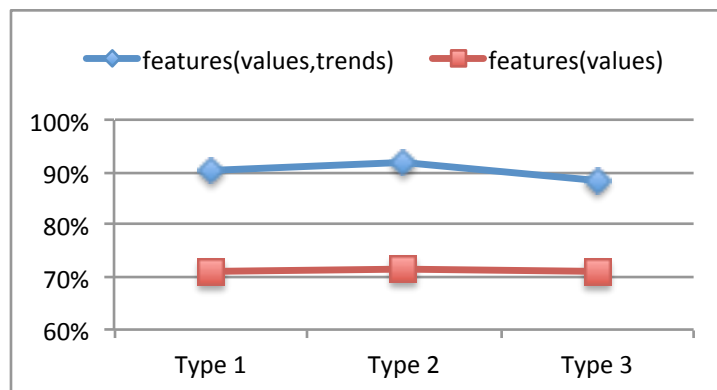


Fig. 5. Prediction accuracy improved by adding trend features

In our future work, we will first try to locate the most important trend patterns for stroke outcome predictions. Then we will work with healthcare professionals to find clinical ground truth beneath those physiological trend patterns of stroke patients. This will greatly benefit clinical treatments of acute ischemic stroke. We also plan to run clinical trials to validate our prediction methods on other real data sets of stroke patients.

References

- [1] Australian Institute of Health and Welfare.: Australia's health 2006, the tenth biennial health report of the Australian Institute of Health and Welfare. ISBN 1 74024 565 2. 2006
- [2] The World Health Organization MONICA Project (monitoring trends and determinants in cardiovascular disease): a major international collaboration. WHO MONICA Project Principal Investigators. *Journal of Clinical Epidemiology*. 1988;41(2):105-14.
- [3] Ye, L., Keogh, E.: Time series shapelets: a new primitive for data mining. *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, Vol. 22, ACM, Paris, France, pp. 947–956.
- [4] Mueen, A., Keogh, E., Young, N.: Logical-shapelets: an expressive primitive for time series classification. *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, Vol. 22, ACM, San Diego, California, USA, pp. 1154–1162.
- [5] Wong, A.: The Natural History and Determinants of Changes in Physiological Variables after Ischaemic Stroke. Ph.D. Thesis, The University of Queensland, St. Lucia.
- [6] Oliveira-Filho, J., Silva, S.C.S., Trabuco, C.C., Pedreira, B.B., Sousa, E.U., Bacellar, A.: Detrimental effect of blood pressure reduction in the first 24 hours of acute stroke onset. *Neurology*. 61(8), 1047–1051.
- [7] Marti-Fabregas, J., Belvis, R., Guardia, E., Cocho, D., Munoz, J., Marruecos, L., Marti-Vilalta, J.-L.: Prognostic value of Pulsatility Index in Acute Intracerebral Hemorrhage. *Neurology*. 61(8), 1051–1056.

Acute Ischemic Stroke Prediction

- [8] Dawson, S.L., Manktelow, B.N., Robinson, T.G., Panerai, R.B., Potter, J.F.: Which Parameters of Beat-to-Beat Blood Pressure and Variability Best Predict Early Outcome After Acute Ischemic Stroke. *Stroke*. 2000(31), 463–468.
- [9] Ritter, M.A., Kimmeyer, P., Heuschmann, P.U., Dziewas, R., Dittrich, R., Nabavi, D.G., Ringelstein, E.B.: Blood Pressure Threshold Violations in the First 24 Hours After Admission for Acute Stroke: Frequency, Timing, Predictors, and Impact on Clinical Outcome. *Stroke*. 2009(40), 462–468.
- [10] Castillo, J., et al., Blood pressure decrease during the acute phase of ischemic stroke is associated with brain injury and poor stroke outcome. *Stroke*, 2004. 35(2): p.520-6
- [11] Ahmed, N., P. Nasman, and N.G. Wahlgren, Effect of intravenous nimodipine on blood pressure and outcome after acute stroke. *Stroke*, 2000. 31(6): p. 12505.
- [12] Yong, M. and M. Kaste, Association of characteristics of blood pressure profiles and stroke outcomes in the ECASSII trial. *Stroke*, 2008. 39(2): p. 36672
- [13] Wong AA, Schluter PJ, Henderson RD, O'Sullivan JD, Read SJ. The natural history of blood glucose within the first 48 hours after ischemic stroke. *Neurology* 2008;70:103641.
- [14] Christensen, H., A. Fogh Christensen, and G. Boysen, Abnormalities on ECG and telemetry predict stroke outcome at 3 months. *J Neurol Sci*, 2005. 234(12): p. 99103.
- [15] Boysen, G. and H. Christensen, Stroke severity determines body temperature in acute stroke. *Stroke*, 2001. 32 (2): p. 4137.
- [16] Gujjar AR, Sathyaprabha TN, Nagaraja D, Thennarasu K and Pradhan N, Heart rate variability and outcome in acute severe stroke: role of power spectral analysis. *Neurocrit Care*, 2004. 1(3): p. 347-53.
- [17] Christensen, H., A. Fogh Christensen, and G. Boysen, Abnormalities on ECG and telemetry predict stroke outcome at 3 months. *J Neurol Sci*, 2005. 234(1-2): p. 99-103.
- [18] Rankin J (May 1957). Cerebral vascular accidents in patients over the age of 60. II. Prognosis. *Scott Med J* 2 (5): 200-15.