Mining Acute Stroke Patients' Data using Supervised Machine Learning

Ritu Kundu and Toktam Mahmoodi

Department of Informatics,
King's College London
London WC2R 2LS, United Kingdom
{ritu.kundu, toktam.mahmoodi}@kcl.ac.uk

Abstract. Analysis of data for identifying patterns and building models has been used as a strong tool in different domains, including medical domains. In this paper, we analyse the registry of brain stroke patients collected over fifteen years in south London hospitals, known as South London Stroke Register. Our attempt is to identify the similar patterns between patients' background and living conditions, their cognitive ability, the treatments they received, and the speed of their cognitive recovery; based on which most effective treatment can be predicted for new admitted patients. We designed a novel strategy which takes into account two different approaches. First is to predict, for each of the potential intervention treatments, whether that particular treatment would lead to recovery of a new patient or not. Second is to suggest a treatment (treatments) for the patient based on those that were given to the patients who have recovered and are most similar to the new patient. We built different classifiers using various state of the art machine learning algorithms. These algorithms were evaluated and compared based on three performance metrics, defined in this paper. Given that time is very crucial for stroke patients, main motivation of this research work is identifying the most effective treatment immediately for a new patient, and potentially increase the probability of their cognitive recovery.

Keywords: Data Mining, Modelling and Analysis of Clinical Data, Machine Learning Algorithms

1 Introduction

Brain stroke is one of the major health concerns. According to the statistics [1], in 2010 stroke was the fourth largest cause of death in the UK after cancer, heart disease and respiratory disease; causing almost 50,000 deaths. Further, more than half of all stroke survivors are left dependent on others for everyday activities. There are approximately 152,000 brain strokes reported in the UK every year, and it is a leading cause of adult disability. According to a recent report, there is an alarming increase in the numbers of people having a stroke in working age [2].

It is important to treat stroke patients immediately with the most efficient treatment. Understanding how recovery and treatments are influenced by patients' *individual* extent of injury (brain damage), and their socio-demographic and medical background could results in more faster and more effective treatments. In other words, individual stroke treatment decision making is more probable to be successful. Prognosis for recovery of an acute stroke patient given a particular intervention treatment can aid the decision making by healthcare professionals. Moreover, suggestions of potentially effective treatments for a new patient based on the other patients with similar clinical, medical and socio-demographical factors who have recovered in the past, can be a guide to decide an appropriate and reliable treatment approach. Prognosis can be done using a model that is based on classifiers which use a set of pre-treatment assessment variables for prediction (classification). These classifiers can be built using machine learning techniques as an alternative to the usual approach of analysing the stroke-data through logistic regression models. Machine learning approach allows exploration of the data leading to interesting, previously unknown, patterns being revealed. Additionally, the greatest strength of machine learning techniques lies in their potential to improve performance by easily incorporating newly available data [24].

Presented here, is an observational study that explores the possibility of using various machine learning techniques to build a tool for assisting medical experts in selecting the most effective intervention-treatment approach. For cases such as stroke, there is a very small window of time, during which treatments can be the most effective and hence a fast and accurate choice of treatment can significantly increases the chances of recovery for a patient after an acute stroke. More specifically, we study various machine-learning algorithms that can be used to train distinct classifiers which can, later, be combined into one software tool, or application, to be used by medical experts.

The paper is organised as follows: The next section provides a short literature review and specifies the contribution of this study. The following section gives an insight into the data used for this study, followed by Section 4 which explains the details of the methodology adopted. After that, Section 5 covering the aspects of modelling the data follows. Results obtained are presented and discussed in Section 6. Section 7 concludes the study presented.

2 Related Work

Thus far, there have been a few studies adopting machine learning techniques for analysing the stroke-data. Recently, a study utilizes one machine learning technique, support vector machine (SVM), on computerized tomography (CT) images along with clinical variables, for prediction of symptomatic intracranial haemorrhage (SICH) associated with intravenous thrombolysis administered to acute ischemic stroke patients [6]. Another recent study applies machine learning algorithms in acute ischemic stroke outcome prediction in relation to treatment by endovascular intervention [4]. Another study, using MRI of rats, compares five predictive algorithms (generalized linear model (GLM), generalized additive model, support vector machine, adaptive boosting, and random forest) to predict brain infarction, and differentiate potentially salvageable tissue from irreversibly damaged tissue [7]. Yet another study uses spatially regularized SVM on brain images of acute stroke patients to detect brain areas associated with motor outcome at 90 days, based on diffusion-weighted images acquired at the acute stage [9].

Our Contribution There are numerous instances of the research work done earlier which employ machine learning techniques for predicting the outcomes of the stroke treatment. However, the prognostic models designed in these studies only focus on whether administering a particular treatment will be beneficial; moreover, only one machine learning method (SVM) is mostly used.

To the best of our knowledge, no attempts have been made to apply a broad range of machine learning algorithms for building a *comprehensive multifactorial model* which can predict the outcome of each of the usual intervention-treatment approaches and can suggest, in addition, potentially more promising treatments based on the *similarity of the patient* to the patients who have recovered. More specifically, this is the first study which incorporates the following:

1. Applying a range of machine learning algorithms for building a comprehensive multifactorial model which can predict the outcome of each of the usual intervention-treatment approaches.
2. Devicing a strategy that is working on two horizons:
   - First Approach: Predicting whether a patient would recover or not if a particular treatment is used, for each of the possible treatments.
   - Second Approach: Suggesting the subtype of each treatment-type based on similarity of the patient with recovered patients.

3 Dataset

The dataset used in the study is a sample set obtained from the community-based South London Stroke Register (SLSR) [1]. The SLSR is a prospective population-based stroke register set up in January 1995, recording all first-ever strokes in patients of all ages for an inner area of South London based on 22 electoral wards in Lambeth and Southwark, over 20 years. In the data-set being used, patients were assessed for cognitive function using Abbreviated Mental Test [11] or Mini-Mental State Examination [15] at the onset, 3 months, and annually thereafter. In addition, various details related to socio-demographics, various risk factors prior to stroke, previous medical history, stroke symptoms and the severity of stroke are noted and taken into account to decide for the acute intervention methods. Later, various medical tests like ECG, ECHO, blood investigations and brain imaging are used for stroke classification.

[1] details can be found online at: KCL Faculty of Life Sciences and Medicine, Stroke research group

4 Methodology

Our methodology to study the SLSR dataset comprised of three steps, as follows:

1. We designed the strategy for addressing the problem, i.e. defined the criteria to build classifiers and identified the classes to train classifiers.
2. Data was pre-processed which included classifying the data, cleaning up the data, bringing it in the form required by the classification techniques and splitting the dataset to obtain the required subsets.
3. Data was modelled by training classifiers on the pre-processed data using different classification algorithms. Training was followed by comparing the performance of the different classification algorithms to study the pros and cons of various classifiers with respect to their application to the SLSR dataset.

The first two steps have been described in the following subsections while the third step has been elaborated in Section 5.

4.1 Defining Classifications Criteria

Figure 1 demonstrates different types (classes) of treatments and the sub-types under each main class. The solution was designed using two different strategies for developing a prediction-model that could assist in choosing the most promising intervention-treatment.

– **First Approach: Predicting whether a patient would recover or not if a particular treatment is used.**
  In this approach each sub-type of the main treatment types was viewed as representation of a *treatment-class*. For each type of the possible treatment-classes, a classifier was built to predict whether that particular treatment-type would result in a patient being recovered or not.
– **Second Approach: Suggesting the subtype of each treatment-type based on similar records of recovered patients.**
  Each record of a recovered patient, in this approach, was seen as a point in a space with number of dimensions equal to that of predictive-attributes being considered. For prediction, similarity of the new record (new patient) from the other records (patients who recovered in the past) was calculated and 'k' (a pre-defined positive constant number) nearest neighbours were decided to be considered. Class of the majority of these k nearest neighbours, i.e. treatment given to the majority of these *similar patients*, would be assigned to the new record. Thus, for each treatment type, its subtype could be suggested using this approach.

4.2 Pre-processing Data

For training classifiers, instances (records) had to be dichotomised in classes corresponding to recovered and not-recovered patients. It was decided after a

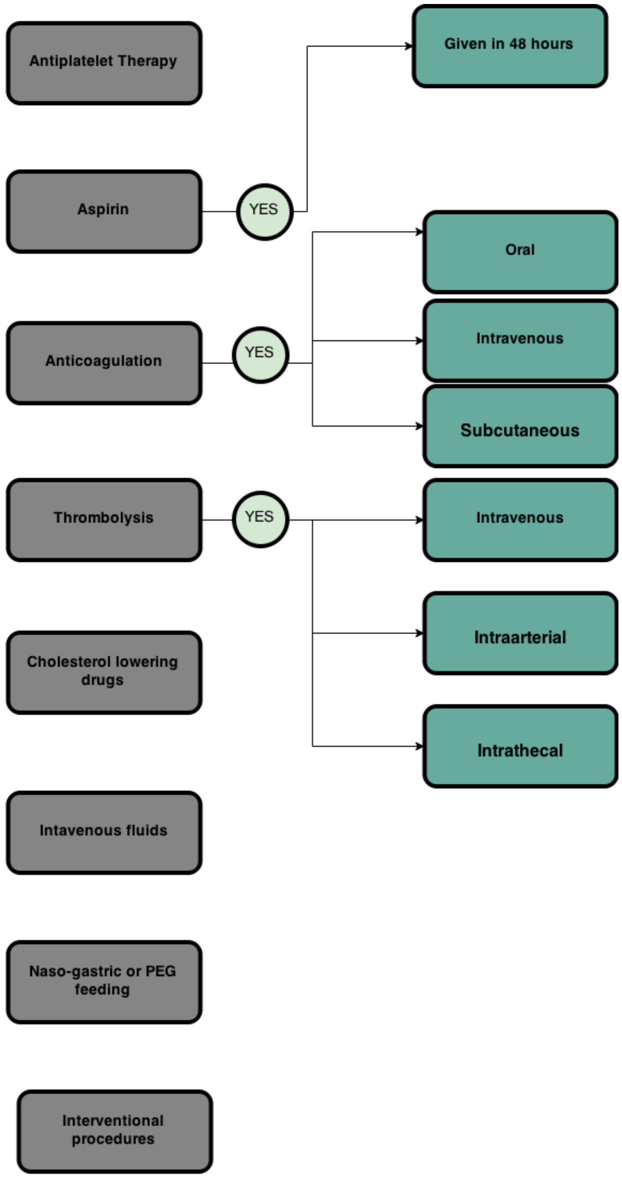


Fig. 1: Different types of acute intervention treatments and their sub-types

discussion with a medical expert that the recovery status could be determined by the scores used to access cognitive impairment. In the given dataset, following scores were recorded:

Table 1: Different types of treatments considered in the classifiers.

Treatment index	type of treatment
1	Antiplatelet therapy (marked as Antiplatelet)
2	Aspirin
3	Anticoagulation-subcutaneous (marked as Anticoag subcutaneous)
4	Anticoagulation-oral (marked as Anticoagulation oral)
5	Thrombolysis-oral (marked as Thrombo oral)
6	Cholesterol lowering drugs (marked as Cholesterol)
7	Naso-gastric or PEG feeding (marked as NP feeding)
8	Intavenous fluids (marked as Intavenous)

– AMT score (recorded 1 January, 2000 onwards) : Abbreviated mental test score is assigned on a scale of 10. 7 is used as the threshold, i.e. score less than or equal to 7 implies cognitive impairment [25].
– MMSE score (recorded before 1 January, 2000): Mini-mental state examination score is rated on a scale of 30 and 24 is considered as the threshold [25].

Since AMT score was observed in the follow-ups as well, an appropriate criteria seemed to label a record as *recovered* where AMT score was above the threshold for every observation and tag a record as *not-recovered* if AMT score was below the threshold for each observation. In the cases where MMSE score was noted initially, it was scaled accordingly to reflect a corresponding value on AMT scale of 10. Since there was no clear monotonous trend in the observations, moving average technique was used to smooth out the short-term fluctuations and capture a long-term stable trend. A window of 3 was taken for the running average.

After labelling, cleaning-up was performed followed by categorisation of the numeric values so as to convert them to nominal (as required by many classifier-training algorithms) and removal of unnecessary attributes. We obtained 520 labelled records in total after labelling and cleaning-up.

Next, data was split as follows: For the first approach, data ($n = 520$) was split along the lines of the treatment-classes, i.e. data-records corresponding to the patients given that treatment were separated from the rest. These subsets of the data were then used (one at a time) to train different classifiers for classifying new records into classes corresponding to 'recovered-patients' and 'not-recovered-patient'. All the treatments listed in Table 1 had been considered for the experiment as they had sufficient data to result in an effective classifier.

For the second approach, only the records labelled as 'recovered' were considered ($n = 390$). Subsets of the data had been generated such that the considered category of treatments were non-empty. Therefore, the first seven categories of Table 1 were considered while lines 3 and 4 of the table were considered as a single category of "AntiCogulation". This completes the pre-processing stage of our methodology.

5 Modelling Data

For modelling SLSR data, we used a number of supervised-learning algorithms representing different approaches of machine learning and these also have been used widely in data-mining studies done so far on medical data. After training the classifiers, their performance in modelling the data was examined and compared.

5.1 Software Tool

The software tool used for training the classifiers as well as for evaluating their performance is *Weka* [13]. Weka (Waikato Environment for Knowledge Analysis), developed by the University of Waikato, is a cross-platform open source and one of the most popular software for machine learning based applications. It has been written in Java and contains a collection of visualization tools and algorithms for data-mining tasks such as data analysis and predictive modelling [2].

5.2 Classification Techniques

In principle, any of the machine learning algorithms can be used to train a classifier, but each of them has its own benefits and limitations depending on the type of data it is being applied. Following algorithms were employed based on their default parameters of the WEKA application on the subsets of the data obtained by splitting the pre-processed and cleaned data.

First Approach

1. Naive Bayes classifier: Probabilistic classifier based on Bayes Theorem. For each class value, it predicts the probability that a given instance belongs to that class. The class having the highest probability is assigned to that instance [16].
2. Support Vector Machine (SVM): It is hyper-plane classifier based on margin maximisation between the target classes by mapping input space to higher dimensional space and thus achieving linear separability [17].
3. Multilayer Perceptron (MLP): A feed-forward back-propagation network consisting of input, output, and one or more hidden layers. It extracts useful information while learning to assign weighted coefficients to components of the input layer [22].
4. Conjective Rule based classifier: It is based on a decision-making rule that uses AND logical relation to correlate stimulus attributes. This rule consists of antecedents (attributes)"AND"ed together and the class value for the classification. It uses *Information Gain* to select the antecedent [14].

[2] For this study Weka 3.6.11 has been used

5. Decision Tables based classifier: It is also a rule-based classifier that builds a decision table based on the labelled instances [18].
6. Alternating Decision Tree: It is a generalisation of decision tree, voted decision tree, and voted decision stumps. It has two types of nodes - decision node (containing a predicate) and prediction node (containing a single number). It is different from the other decision trees such as C4.5 in which an instance travels only a single path through the tree. In ADTree, an instance follows all the paths for which decision nodes are true, summing up any predicate nodes traversed [12].
7. C4.8: It is an extension of Quinlan's earlier ID3 algorithm. It produces a decision tree from labelled instances using the concept of information entropy [21]. It is 'J48' in Weka.
8. Naive Bayes classifier based Decision Tree : It is a hybrid of decision tree and naive Bayes classifiers. It produces a decision tree with naive Bayes classifiers at the leaves [20].

Second Approach

1. K-nearest neighbours classifier(KNN): It is an instance based classifier that uses similarity of a given instance with other instances to choose its neighbours and uses the majority of neighbours to classify that instance. K is a positive integer [3]. It is same as 'IBk' in Weka.
2. KStar (K*): It is also an instance-based classifier similar to KNN classifier but it uses entropy as the distance measure [8].

The value of $k = 3$ has been used for the above two algorithms.

5.3 Testing Technique

10-fold Cross Validation [23] was used as the testing technique to minimize the bias associated with random sampling of training and test data samples [19].

5.4 Performance Metrics for Evaluation

We have evaluated different classification techniques using three performance indicators: Prediction accuracy, Kappa measure, and Area under Receiver Operating Characteristics (ROC) curve.

The prediction accuracy was computed based on the proportion of instances classified correctly. In this study, accuracy was calculated by averaging the results from 10 runs of 10-fold cross validation. The second performance metric chosen was Kappa measure because Cohen's Kappa statistic is mostly in agreement with the overall accuracy but proves to be better in case of unbalanced datasets as it compensates for the classification that may be due to chance [5]. The third metric selected was Area under ROC Curve (AUC). A Receiver Operating Characteristics (ROC) graph is a technique for selecting classifiers based on their performance. It is one of the most commonly used evaluation criteria of classifiers in medical-domain. Area under the two-dimensional ROC curve is a well-used method for comparing classifiers [10].

6 Evaluation Results and Comparison

Figures 2a, 2b, and 2c demonstrate evaluation results from the first approach, corresponding to the chosen performance metrics (Accuracy, Kappa measure and AUC). Evaluation results obtained for the second approach are illustrated in the Figures 3a, 3b, and 3c. All of these figures also compare different algorithms used for training each of the classifiers (corresponding to every data subset). These results achieved by evaluating various classifiers are analysed and discussed in the following subsections.

6.1 First Approach

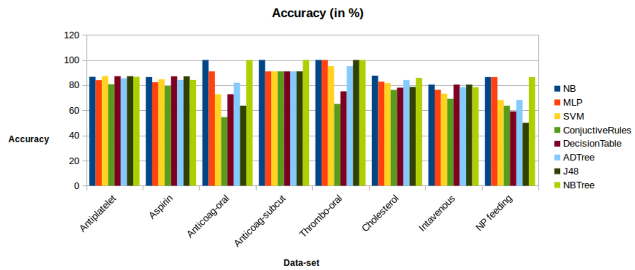
We summarize the results from the first approach (Figure 2) in Table 2, in which the best performing algorithms based on each of the performance-metrics on each of the data-subsets can be seen. Overall, NB, MLP and tree-based algorithms are performing quite well, in comparison with the rule-based algorithms. Additionally, it can be seen that Accuracy and Kappa measure are in agreement with each other, but AUC has different behaviour for a particular dataset. Moreover, it can be observed that SVM is performing worst on all of the datasets according to all three comparison-parameters.

Table 2: First approach: Best performing classification algorithms for each of the data-subsets

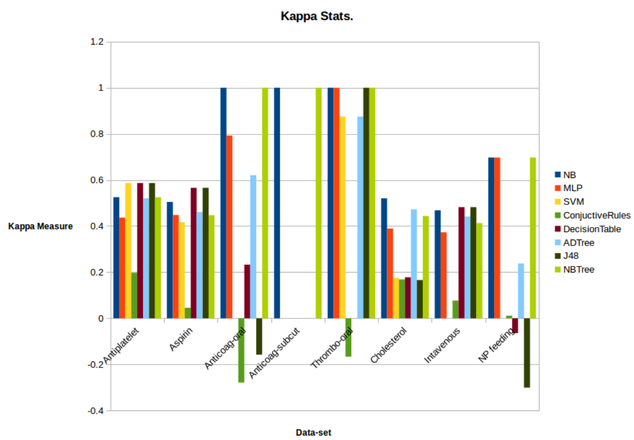
	Accuracy (in %)	Kappa measure	Area under ROC
Antiplatelet	J48	J48	ADTree
	87.0968	0.578	0.871
Aspirin	ADTree	ADTree	NB
	85.7988	0.5375	0.846
Anticoag-subcut	All except SVM	NB, NBTree	NB, NBTree
	90.9091	0.6207	1
Anticoag-oral	MLP, ConjuctiveRules, ADTree, J48	NB, DecisionTable, NBTree	ADTree
	72.7273	0.2326	0.75
Thrombo-oral	NB, J48, NBTree	J48	NB, J48, NBTree
	90	0.76447	1
Cholesterol	ADTree	ADTree	ADTree, J48
	86.3095	0.4889	0.883
Intavenous	ADTree	ADTree	ADTree
	79.3814	0.4476	0.807
NP feeding	MLP	MLP	NB, NBTree
	81.8182	0.581	0.867

6.2 Second Approach

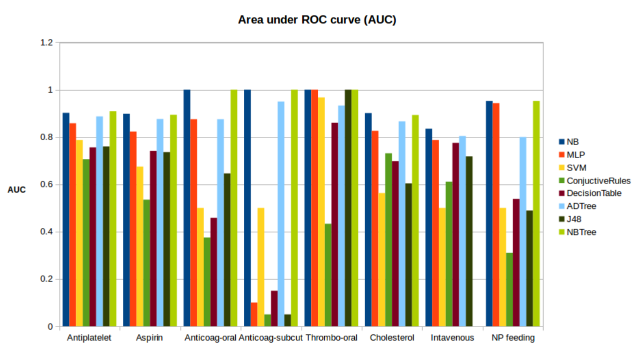
As can be seen from Table 3, both of the chosen algorithms are performing bad on the basis of Kappa measure. However, based on the other two metrics, it can



(a)

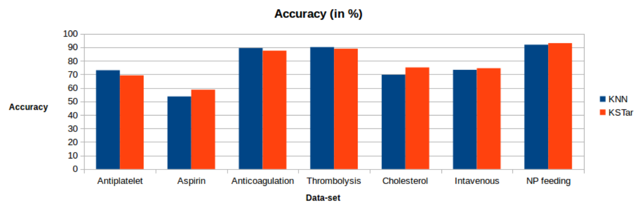


(b)

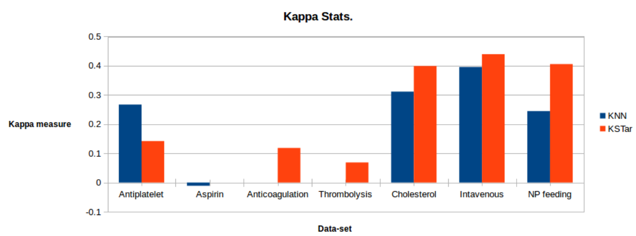


(c)

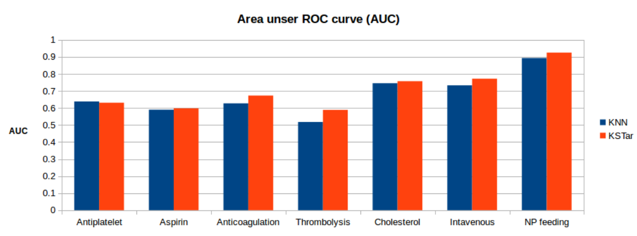
Fig. 2: First approach: Comparison of classifiers' (a) accuracy (b) Kappa measure (c) AUC.



(a)



(b)



(c)

Fig. 3: Second approach: Comparison of classifiers' (a) accuracy (b) Kappa measure (c) AUC.

be inferred that KStar is performing better than KNN for the first four datasets while KNN is the clear winner for the remainder of three datasets. In addition, it seems that all the three comparison parameters are approximately in agreement with each other for this approach, which is in contrast to the results obtained for the first approach.

Table 3: Second approach: Best performing classification algorithms for each of the data-subsets

	Accuracy (in %)	Kappa measure	Area under ROC
Antiplatelet	Both	KStar	KStar
	65.019	0.168	0.612
Aspirin	KStar	KStar	KStar
	54.5455	0.2287	0.634
Anticoagulation	KNN	KStar	KStar
	88.9734	0.0504	0.565
Thrombolysis	KNN	KNN	KStar
	88.5496	0.0041	0.562
Cholesterol-oral	KNN	KNN	KNN
	64.9805	0.1647	0.68
Intavenous	KNN	KNN	KNN
	67.6113	0.2676	0.709
NP feeding	KNN	KStar	KNN
	91.8605	0.179	0.82

7 Conclusion

In this paper, we applied data mining methods, i.e. machine learning, to the historical data from stroke patients obtained from SLSR. This dataset recorded various details of the patients related to socio-demographics, various risk factors prior to stroke, previous medical history, stroke symptoms, and the severity of stroke. We designed a novel strategy which takes into account two different approaches. First, to predict, for each of the potential intervention treatments, whether that particular treatment would lead to recovery of a new patient or not. Second, to suggest treatment(s) for the new patient, based on those which were given to the patients in the past, with similar profile to the new patient, and were successful. We built different classifiers using various machine learning algorithms. These algorithms were evaluated and compared based on three performance metrics: Accuracy, Kappa measure, and Area under ROC. Comparison of the algorithms for each of the classifiers let us gain a clearer insight into which classification algorithm would work better on which of the classifier (subset of the dataset).

Given that most of the existing studies on medical data using machine learning techniques are limited in terms of focussing on prediction of the outcome after stroke for a particular treatment, there is a need for more thorough models. Our proposed model, therefore, covers all the possible treatment options at the same time and results in more accurate analysis, if used with a sufficiently large dataset. The outcome of our model can be used in clinical decision making to significantly increase the probability of recovery by choosing the appropriate intervention treatment.

The ultimate, more usable, outcome of this work can be in the form of a user friendly software that combines different classifiers from both the approaches which are analysed here, for making the treatment suggestions. This software application can be used by medical experts to assist them in quickly choosing the most promising treatment so that the chances of recovery of the patient increases.

8 Acknowledgement

The authors would like to thank Miss Siobhan Crichton from the department of Primary Care & Public Health Sciences at King's College for providing support on the dataset, to the authors. Also, we would like to thank Dr. Arshia Sedigh at King's College Hospital for his valuable advice on understanding the data without which this work was not possible.

References

1. State of the Nation: stroke statistics. Stroke Association (January 2015), `http://www.stroke.org.uk/`
2. Strokes rising among people of working age, warns charity. BBC Health news (March 2015), `http://www.bbc.com/news/health-32690040`
3. Aha, D., Kibler, D.: Instance-based learning algorithms. Machine Learning 6, 37–66 (1991)
4. Asadi, H., Dowling, R., Yan, B., Mitchell, P.: Machine learning for outcome prediction of acute ischemic stroke post intra-arterial therapy. PLoS ONE 9 (February 2014), `http://dx.doi.org/10.1371%2Fjournal.pone.0088225`
5. Ben-David, A.: Comparison of classification accuracy using cohen's weighted kappa. Expert Systems with Applications 34(2), 82–832 (2009)
6. Bentley, P., Ganesalingam, J., Jones, A.L.C., Mahady, K., Epton, S., Rinne, P., Sharma, P., Halse, O., Mehta, A., Rueckert, D.: Prediction of stroke thrombolysis outcome using CT brain machine learning. NeuroImage: Clinical 4, 635–640 (2014)
7. Bouts, M.J., Tiebosch, I.A., van der Toorn, A., Viergever, M.A., Wu, O., Dijkhuizen, R.M.: Early identification of potentially salvageable tissue with mri-based predictive algorithms after experimental ischemic stroke. Journal of Cerebral Blood Flow & Metabolism 33(7), 1075–1082 (2013)
8. Cleary, J.G., Trigg, L.E.: K*: An instance-based learner using an entropic distance measure. In: 12th International Conference on Machine Learning. pp. 108–114 (1995)

9. Cuingnet, R., Rosso, C., Lehéricy, S., Dormont, D., Benali, H., Samson, Y., Colliot, O.: Spatially regularized svm for the detection of brain areas associated with stroke outcome. In: Medical Image Computing and Computer-Assisted Intervention (MICCAI), pp. 316–323. Springer (2010)
10. Fawcett, T.: An introduction to roc analysis. Pattern Recognition Letters 14(27), 861–874 (2006)
11. Folstein, M.F., Folstein, S.E., McHugh, P.R.: mini-mental state: a practical method for grading the cognitive state of patients for the clinician. Journal of psychiatric research 12(3), 189–198 (1975)
12. Freund, Y., Mason, L.: The alternating decision tree learning algorithm. In: Proceeding of the Sixteenth International Conference on Machine Learning. pp. 124–133. Bled, Slovenia (1999)
13. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H.: The weka data mining software: An update. SIGKDD Explorations 11(1), 385–403 (2009)
14. Har-Peled S., Roth D., Z.D.: Constraint Classification for Multiclass Classification and Ranking. In: 16th Annual Conference on Neural Information Processing Systems, NIPS-02. pp. 785–792. MIT Press (2003)
15. HODKINSON, H.M.: Evaluation of a mental test score for assessment of mental impairment in the elderly. Age and Ageing 1(4), 233–238 (1972)
16. John, G.H., Langley, P.: Estimating continuous distributions in bayesian classifiers. In: Eleventh Conference on Uncertainty in Artificial Intelligence. pp. 338–345. Morgan Kaufmann, San Mateo (1995)
17. Kecman, V.: Learning and Soft Computing Support Vector Machines, Neural Networks, Fuzzy Logic Systems. The MIT Press (2001)
18. Kohavi, R.: The power of decision tables. In: 8th European Conference on Machine Learning. pp. 174–189. Springer (1995)
19. Kohavi, R.: A study of cross-validation and bootstrap for accuracy estimation and model selection. Proceeding of 14th International Joint Conference on Artificial Intelligence, (IJCAI'95) pp. 1137–1143 (1995)
20. Kohavi, R.: Scaling up the accuracy of naive-bayes classifiers: A decision-tree hybrid. In: Second International Conference on Knoledge Discovery and Data Mining. pp. 202–207 (1996)
21. Quinlan, R.: C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers, San Mateo, CA (1993)
22. Rumelhart, D.E., Hinton, G.E., Williams, R.J.: Parallel distributed processing: Explorations in the microstructure of cognition,. vol. 1, chap. Learning Internal Representations by Error Propagation, pp. 318–362. MIT Press (1986)
23. Varma S, S.R.: Bias in error estimation when using cross-validation for model selection. BMC Bioinformatics 7(91) (2006)
24. Warwick, K.: March of the Machines: The Breakthrough in Artificial Intelligence. University of Illinois Press (2004)
25. Wolfe, C.D., Crichton, S.L., Heuschmann, P.U., McKevitt, C.J., Toschke, A.M., Grieve, A.P., Rudd, A.G.: Estimates of outcomes up to ten years after stroke: analysis from the prospective south london stroke register. PLoS medicine 8(5) (2011)