Building a Classifier of Onset Stroke Prediction Using Random Tree Algorithm

Yu-Chen Chen, Takashi Suzuki, Masaaki Suzuki, Hiroyuki Takao, Yuichi Murayama, and Hayato Ohwada

Abstract—**A stroke is the result of cell death caused by poor blood flow or vascular obstruction in the brain, but it normally happens suddenly and is hard to prevent. In addition, strokes are one of the main causes of death, with many people dying from this disease every year. Hence, using the latest technology to predict strokes is an important concern.**

In this study, we built a classifier based on the simulation and practical data both in 50 data sets from Japanese patients. We used attribute selection (CfsSubsetEval evaluator and Greedy Stepwise method) and Decision Trees (Random Tree algorithm) to build the classifier with high accuracy. In the end, the result demonstrates that our method, using data that merges the simulation and practical data, can achieve high accuracy.

Index Terms—**Stroke prediction, random tree algorithm, WEKA, select attributes, machine learning.**

I. INTRODUCTION

A stroke is the result of cell death caused by poor blood flow or vascular obstruction in the brain, but it normally happens suddenly and is hard to prevent. In Japan, stroke ranks third among all causes of death, following cancer and heart disease [1]. As of 2014, an average of 130,000 Japanese people die of stroke per year, and there are over one million Japanese people suffering from cerebrovascular diseases which are potential factors for stroke. Patients' annual medical expenses reached 1821 hundred million Japanese yen in 2014 [2]. In addition, according to trends in Japan's aging population, the population of Japan will reach 117 million in 2030 [3]. This means that the number of stroke patients and their medical expenses will grow. For these reasons, the issue of predicting and treating stroke should not be ignored.

At the same time, following the innovative technology developed in recent years, the predicted result with high accuracy has become possible. Machine learning is one of the popular technologies of recent years, as it helps researchers that can make decisions or predictions by building a model from sample inputs. Especially, using different algorithms can produce different performance results. We applied Machine learning technologies in our study and developed a classifier with high accuracy. We expect that the result will help a lot of doctors to issue objective diagnoses and will also help patients get the correct treatment earlier. In the near future, stroke prediction will lead to a decrease in the number of patients, a reduction in medical expenses, and will even benefit Japan itself and its aging society.

Therefore, the objective of this study was to build a classifier for onset stroke prediction using simulation and practical data. Our classifier includes a total of 49 features using a Random Tree algorithm. Data were obtained for 46 patients whose aneurysm did not rupture and six patients whose aneurysm did rupture. We processed the data into a form that WEKA software can understand then applied it to WEKA and built a classifier to predict a ruptured aneurysm. The result was a good performance classifier, so when doctors issue diagnoses based on the use of our classifier, they can successfully predict the onset of strokes.

II. MATERIALS AND METHODS

This study used 50 data sets provided by the Jikei University School of Medicine (JUSM). These samples have 49 attributes based on simulations of blood flow using computational fluid dynamics (CFD) data and medical history data (MH) data. Fig. 1 is a flow diagram of this study that considers the CFD data, MH data, and merged data (i.e., CFD data with MH data) together. We will provide a further explanation of these data below.
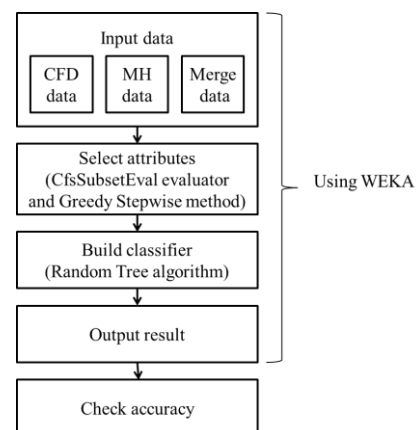
Y. C. Chen is now with the Tokyo University of Science, Noda City, 278-8510 Japan (e-mail: 7416647@ed.tus.ac.jp).

T. Suzuki is now with the Tokyo University of Science, Katsushika-ku, 125-8585 Japan (e-mail: 4514704@ed.tus.ac.jp).

M. Suzuki and H. Ohwada are now with the Department of Industrial Administration, Tokyo University of Science, Noda City, 278-8510 Japan (e-mail: m-suzuki@rs.tus.ac.jp, ohwada@rs.tus.ac.jp).

H. Takao is now with the Department of Neurosurgery and Innovation for Medical Informaion Techology, Jikei University School of Medicine, Tokyo, 105-8461 Japan (e-mail: takao@jikei.ac.jp).

Y. Murayama is now with the Department of Neurosurgery, Jikei University School of Medicine, Tokyo, 105-8461 Japan (e-mail: ymurayama@jikei.ac.jp).

Fig. 1. Flow Diagram of This Study.

In building a highly accurate classifier for stroke prediction, this work used Machine learning techniques, an attribute subset evaluator (CfsSubsetEval), a search method (Greedy Stepwise), and Decision Trees (i.e., a Random Tree algorithm), by WEKA.

A. Stroke Prediction

Stroke a major cause of death and serious long-term disability. A stroke can affect a person at any age, and it usually occurs suddenly. Actually, strokes can be prevented through therapeutic manipulation, and the modifiable risk factors are crucial [4]. Today there is much collected data regarding cases of various diseases in medical science [5]. However, a patient who suffers a stroke usually depends on the doctor's medical knowledge for a diagnosis. Therefore, as modern medicine and innovative technologies have been developed in recent years, they enable predicting strokes with high accuracy. For example, the development of simulated blood vessels provides us a different way of thinking about the causes of diseases. That is, the predicted result can help doctors to objectively diagnose, and doctors can provide the appropriate treatment to patients early. For these reasons, we were motivated to complete this study.

B. Computational Fluid Dynamics Simulations of Blood Flow (CFD) Data

As technology has progressed, image-based computational fluid dynamics has been widely used to obtain hemodynamic fields [6], [7]. Consequently, our cooperative research unit, JUSM, simulated the patients' blood vessels as a 3D shape and collected the relevant information. The unit then identified seven physical blood-flow characteristics, including the pressure loss coefficient (PLC), the energy loss (EL), the exposure limit value (ELV), the maximum wall shear stress (WSS), the average WSS, the minimum WSS, and the oscillation index (OSI). Furthermore, this study considers the maximum, minimum, amplitude, and average of these physical quantities. In all, there are a total of 28 attributes in the CFD data that were available for use in this study.

C. Medical History (MH) Data

The medical history (MH) data are related to stroke incidence and type and to survival and recurrence following the stroke [8]. Thus, in addition to using the CFD data, this study also considered patients' MH data. There are two types of MH data. The first is categorical data. This contains the patient's gender and history of subarachnoid hemorrhage (SAH), smoking, blebs, diabetes mellitus (DM), hypertension (HT), hyperlipidemia, alcohol consumption, polycystic kidneys, cerebral infant, cerebral hemorrhage, past history and hormone replacement, as well as the patient's family history of SAH, unruptured aneurysm, and polycystic kidneys. The other type is continuous data, including the size of the aneurysm, the final age of consultation, the first length of the aneurysm, the maximum size of the aneurysm, and the first neck of the aneurysm. Overall, the MH data provides 21 attributes.

D. WEKA

The Waikato Environment for Knowledge Analysis (WEKA) is a software package that collects Machine learning algorithms and data preprocessing tools [9]. It provides a unified workbench that allows researchers easy access to state-of-the-art techniques [10]. This study used the latest version of WEKA (3.8). Because of its various advantages, it is well-suited for developing new Machine learning schemes [11].

E. Attribute Selection

Feature selection is normally done by searching the space of attribute subsets and evaluating each one. In WEKA, the user chooses an attribute subset evaluator and a search method to achieve the selection, which can be done using a full training set or by cross-validation [12]. In this study, we chose the CfsSubsetEval as our attribute subset evaluator and the Greedy Stepwise search method, using cross-validation to build our classifier.

1) Attribute Subset Evaluator (CfsSubsetEval)

CfsSubsetEval is an attribute subset evaluator that considers the predictive value of each attribute individually, along with the degree of redundancy among them. CfsSubsetEval evaluates the value of a subset of attributes by considering the individual predictive ability of each feature along with the degree of redundancy among them, preferring subsets of attributes that are highly correlated with the class while having low inter-correlation [9], [12].

For eliminating redundant and irrelevant attributes, CfsSubsetEval selects subsets of attributes that individually correlate well with the class while having little inter-correlation.

Consequently, we use the symmetric uncertainty, calculated by Eq. (1), to measuring the correlation between two nominal attributes A and B.

$$\mathrm{U}(A,B) = 2\frac{H(A) + H(B) - H(A,B)}{H(A) + H(B)} \tag{1}$$

In Eq. (1), we use the entropy function from Eq. (2) to calculate H, where p represents the property in the entropy formula, Eq. (2).

$$\begin{aligned} H(p_1, p_2, ..., p_n) = \\ - p_1 \log p_1 - p_2 \log p_2 ... - p_n \log p_n \end{aligned} \tag{2}$$

In addition, $H(A,B)$ represents the joint entropy of A and B, which is calculated based on the joint probabilities of all combinations of values of A and B. According to Eq. (3), we could obtain the "goodness" of a set of attributes using correlation-based feature selection.

$$\frac{\sum_j U(A_j, C)}{\sqrt{\sum_i \sum_j U(A_i, A_j)}} \tag{3}$$

Within Eq. (3), C represents the class attribute and the indices i and j range over all of the attributes in the set [9].

2) Search Method (Greedy Stepwise)

The Greedy Stepwise is one of search methods that greedy

hill-climbing without backtracking and optionally generates a ranked list of attributes. This method searches "greedily" through the space of attribute subsets; it may progress forward from the empty set or backward from the full set. It does not backtrack but terminates as soon as adding or deleting the best-remaining attribute decreases the evaluation metric [9].

This search method identified the subset with the highest merit in terms of explaining the toxicity, determined by empirical changing of variables [13]. It considers both adding and removing features at each decision point, which allows retraction of an earlier decision without keeping explicit track of the search path. After attributes are generated, one can select the best or simply the first attribute that improves accuracy over the current set. In an alternative mode, the method ranks the attributes by traversing the space from empty to full (or vice versa) and recording the order in which attributes are selected [14].

F. Classifier Building

The classify function in WEKAS provides numerous categories to choose, such as Bayesian Classifiers, Generating Rules, and Decision Trees. Decision Trees is one of the more popular classification categories, because Decision Trees represent a supervised approach to classification that utilizes a simple structure wherein non-terminal nodes represent tests on one or more attributes and terminal nodes reflect decision outcomes [15].

The Decision Trees category has many algorithms for the user to choosing by the user for their request. In this study, we used one of the Decision Tree algorithms called Random Tree with its maximum depth set to three by cross-validation, because Random Tree's performance was found to be better than the C4.5 algorithms and CART (Classification and Regression Tree) algorithms in this study (see Table V in Appendix).

Random Tree is a supervised classifier; each node is split using the best among a subset of predictors randomly chosen at that node. There are k random features at each node. "At random" in this context means that, within the set of trees, each tree has an equal chance of being sampled. A random tree is a tree drawn at random from a set of possible trees. Random trees can be generated efficiently, and a combination of a large set of random trees generally leads to accurate models. Random Tree models have been extensively developed in the field of Machine learning in recent years [16].

G. Accuracy Checking

In this study, an error matrix (see Table I) is used to evaluate the predictive accuracy of the ruptured-unruptured model. It includes True Positives (TP), the number of patients for whom rupture was correctly predicted by the model; False Positives (FP), the number of patients in whom no rupture was found but the model predicted a rupture; False Negatives (FN), the number of patients in whom a rupture was found but the model predicted no rupture; and True Negatives (TN), the number of patients for whom the lack of a rupture was correctly predicted by the model.

To measure the predictive accuracy, we calculated the data from the error matrix using the formulas in Table II. This table also gives the meaning for each rate in this study.

TABLE I: THE ERROR MATRIX FOR EVALUATING THE PREDICTIVE ACCURACY OF THE RUPTURED-UNRUPTURED MODEL

Error matrix		Actual result	
		Ruptured	Unruptured
Predicted result	Ruptured	True Positive	False Positive
	Unruptured	False Negative	True Negative

TABLE II: MEASURES OF PREDICTIVE ACCURACY CALCULATED FROM THE ERROR MATRIX (TABLE I)

Measure	Formula	Meaning
accuracy	$\dfrac{TP+TN}{TP+FP+TN+FN}$	The rate of actual ruptured samples was predicted correctly.
precision	$\dfrac{TP}{TP+FP}$	The rate of actual ruptured samples in the predicted result for ruptures.
recall	$\dfrac{TP}{TP+FN}$	The rate of actual ruptured samples that were predicted ruptured.
F-measure	$\dfrac{2 \times precision \times recall}{precision + recall}$	The harmonic mean of Precision and Recall.

III. RESULTS

A. Attribute Selecting

TABLE III: THREE DATA SETS, WITH SELECTED ATTRIBUTES

CFD data	MH data	Merged data
WSSave_max	Sex	SAH
WSSave_amp	SAH	Smoking
-	smoking	Bleb
-	Bleb	Hyperlipidemia
-	DM	Alcohol consumption
-	Hyperlipidemia	Cerebral hemorrhage
-	Alcohol consumption	Family history _SAH
-	Cerebral hemorrhage	WSSmin_ave
-	Family history _unruptured aneurysm	WSSave_min

To compare the performance of the three data sets (CFD data, MH data, and Merge data), an analysis was executed three times as the function of the attributes selected (the CfsSubsetEval evaluator with the Greedy Stepwise search method) and in WEKA. In this manner, we generated a table of the three data sets with selected attributes (see Table III).

The CFD data is represented by two attributes, WSSave_max and WSSave_amp, from among 28 attributes. The two attributes represent the maximum WSS average and the amplitude of the WSS average. The MH data is represented by nine attributes from among 21 attributes, including the patient's sex, the history of SAH, smoking, bleb, DM, hyperlipidemia, alcohol consumption, cerebral hemorrhage, and the family history of unruptured aneurysm. The Merged data combines the CFD data and the MH data, and represents 9 of the 49 attributes, including the history of SAH, smoking, bleb, hyperlipidemia, alcohol consumption, cerebral hemorrhage, the family history of SAH, WSSmin_ave (the average WSS minimum), and WSSave_min (the minimum of the WSS average).

B. Classifier Building

We selected the attributes of the three data sets and

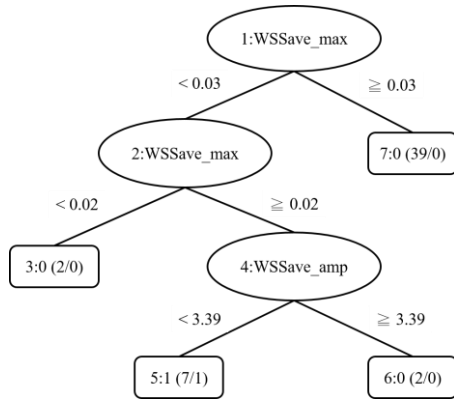individually classified them using the Random Tree algorithm.

Fig. 2. The visualize tree of classified CFD data.

Fig. 2 is the tree drawn by classifying the patient's aneurysm as ruptured with two attributes. According to this result, we assume that WSSave_max and WSSave_amp are the important factors. We also can determine that the aneurysm would have higher rupture probability under the conditions that $0.03>$WSSave_max ≥ 0.02 and WSSave_amp<3.39 in the CFD data.
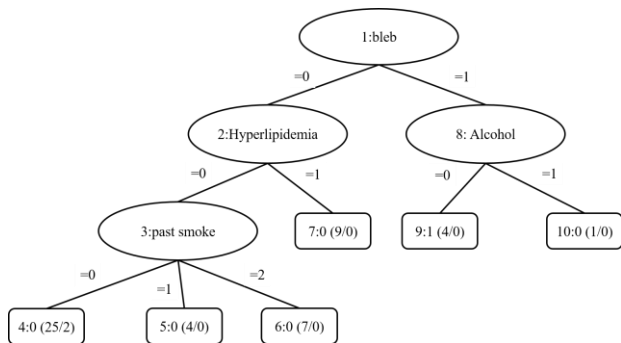

Fig. 3. The visualize tree of classified MH data.

The tree in Fig. 3 was generated by classifying the patient's aneurysm as ruptured using nine attributes in the MH data. As this result, we assume that bleb and alcohol consumption are the important factors. This means that ruptured aneurysm samples appeared under the conditions that the patient has the history of bleb and alcohol consumption.
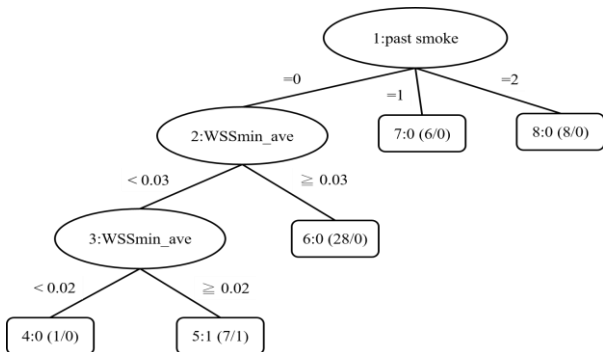

Fig. 4. The visualize tree of classified Merge data.

Fig. 4 is the tree drawn by classifying the patient's aneurysm as ruptured using nine attributes. With this result, we assume that the history of smoking and WSSmin_ave are the important factors, and we can determine that the aneurysm has a higher rupture probability under the conditions that the

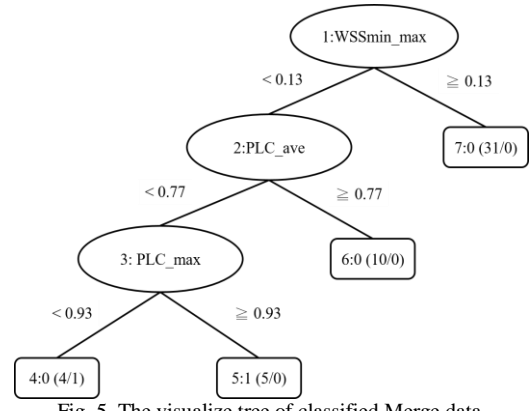patient did not smoke in the past and $0.03>$WSSmin_ave$\geq$ 0.02.

Fig. 5. The visualize tree of classified Merge data.

In addition, we compared the result between the Merge data after selection and the original Merge data. The tree in Fig. 5 was generated by classifying the patient's aneurysm as ruptured using 49 attributes in the MH data. With this result, we assume that WSSmin_max, PLC_ave, and PLC_max are the important factors. This means that ruptured aneurysm samples appeared under the conditions of WSSmin_ave>0.13, PLC_ave<0.77, and PLC_max<0.93.

C. Accuracy Checking

We used the results in the previous paragraph for calculating the accuracy and generated Table IV, which summarizes the accuracy, precision, recall, and F-measure of the four data sets.

TABLE IV: THE TABLE OF THE THREE DATA OF PREDICTIVE ACCURACY

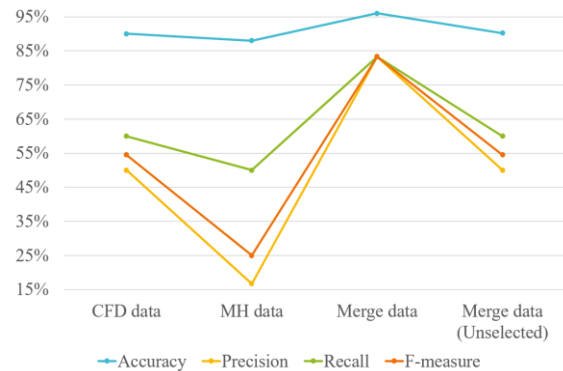| Data	CFD data	MH data	Merge data	Merge data (Unselected)
TP	3	1	5	3
FP	3	5	1	3
FN	2	1	1	2
TN	42	43	43	43
Accuracy	90%	88%	96%	90.2%
Precision	50%	16.7%	83.3%	50.0%
Recall	60%	50%	83.3%	60.0%
F-measure	54.5%	25%	83.3%	54.5%

Fig. 6. Line graph comparing the predictive accuracy of the three data sets.

Fig. 6 is a line graph showing the result of comparing the four data sets. The performance of the selected Merge data set has the highest accuracy, higher than the CFD data, the MH data, and the unselected Merge data. Furthermore, even the precision, recall, and F-measure of the selected Merge data

are better than those of the other three data sets.

IV. CONCLUSION

We referred to background knowledge and examples, built a classifier to predict ruptured aneurysms, and obtained a high-accuracy result that uses Machine learning technologies and combine the simulation and practical data. This study not only successfully built a classifier with very good performance (accuracy 96%, precision 83.3%, recall 83.3%, F-measure: 83.3%) but also identified highly important attributes such as the history of smoking and WSSmin_ave. It showed that patients who were non-smoking in the past and for whom 0.03>WSSmin_ave$\geq$0.02 have a high probability of experiencing a stroke.

Future work includes investigating the stability of the classifier and its applicability to other patients. In this study, we selected 50 patients, but we need to consider a greater number of patients and to incorporate some new features based on experts' suggestions. In addition, we used CFD data in this study, and the simulation data provided the classifier with reference attributes. Future studies will consider the further use of the 3D simulating shape of blood flow and the use of picture recognition technology to make our stroke predictions even more accurate.

We expect that this study and future research will allow more doctors to treat their patients as early as possible and will gradually build a complete system for forecasting strokes in one day, to help people live longer and indirectly reduce medical expenses.

APPENDIX

To decide which method would be used, we compared three different classification algorithms, C4.5, CART, and the Random Tree algorithm with selected Merge data (see Table V).

TABLE V: PREDICTIVE ACCURACY OF THREE CLASSIFICATION ALGORITHMS

Data	C4.5	CART	Random Tree
TP	4	0	5
FP	2	6	1
FN	1	1	1
TN	43	43	43
Accuracy	94.0%	86.0%	96.0%
Precision	66.7%	0.0%	83.3%
Recall	80.0%	0.0%	83.3%
F-measure	72.7%	0.0%	83.3%

ACKNOWLEDGMENT

REFERENCES

[1] K. Kodama, "Stroke trends in Japan," *Annals of Epidemiology*, Richard Rothenberg, vol. 3, pp. 524-528, 1993.

[2] I. Yoko and W. Mika, 2014 *Overview of Patient Survey*, 1st ed. Tokyo, Japan: Ministry of Health, Labour and Welfare, 2015, pp. 3-15.

[3] Japan in 2030: The Future Demographic. (June 2015). *Euromonitor International.* [Online]. Available: http://www.euromonitor.com/japan-in-2030-the-future-demographic/report

[4] A. Sudha, P. Gayathri and N. Jaisankar, "Effective analysis and predictive model of stroke disease using classification methods," *International Journal of Computer Applications*, Amr Ahmed, pp. 26-31, vol. 14, 012.

[5] L. Amini, R. Azarpazhouh, M. T. Farzadfar, S. A. Mousavi, F. Jazaieri, F. Khorvash, R. Norouzi, and N. Toghianfar, "Prediction and control of stroke by data mining," *International Journal of Preventive Medicine*, Mohammad Hassan Emami, pp. 245-249, vol.2, 2013.

[6] J. Xiang, A.H. Siddiqui, and H. Meng, "The effect of inlet waveforms on computational hemodynamics of patient-specific intracranial aneurysms," *Journal of Biomechanics*, Farshid Guilak, pp. 3882-3890, vol. 47, 2014.

[7] J. Xiang, V. M. Tutino, K. V. Snyder, and H. Meng, "CFD: computational fluid dynamics or confounding factor dissemination? The role of hemodynamics in intracranial aneurysm rupture risk assessment," *American Journal of Neuroradiology*, pp. 1849-1857, vol. 10, 2014.

[8] R. L. Sacco, P. A. Wolf, W. B. Kannel, and P. M. McNamara, "Survival and recurrence following stroke — The framingham study," *Stroke*, pp. 290-295, vol. 13, 1982.

[9] I. H. Wittern, E. Frank, and M. A. Hall, *Data Mining: Practical Machine Learning Tools and Techniques*, 3rd ed, San Francisco, U.S.A: Morgan Kaufmann, 2011.

[10] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA data mining software: An update," *SIGKDD Explorations*, pp. 10-18, vol. 11, 2009.

[11] Weka 3: Data Mining Software in Java. (June 2015). *WEKA: The University of Waikato.* [Online]. Available: http://www.cs.waikato.ac.nz/ml/weka/

[12] K. Selvakuberan, M. Indradevi, and R. Rajaram, "Combined Feature Selection and classification - A novel approach for the categorization of web pages," *Journal of Information and Computing Science*, pp. 83-89, vol. 3, 2008.

[13] L. E. Llewellyn, "Predictive toxinology: An initial foray using calculated molecular descriptors to describe toxicity using saxitoxins as a model," *Toxicon*, pp. 901-913, vol. 50, 2007.

[14] R. Sadeghia, R. Zarkamib, K. Sabetraftarb, and P. V. Damme, "Application of genetic algorithm and greedy stepwise to select input variables in classification tree models for the prediction of habitat requirements of Azolla filiculoides (Lam.) in Anzali wetland, Iran," *Ecological Modelling*, pp. 44-43, vol. 251, 2013.

[15] Y. Zhao and Y. Zhang, "Comparison of decision tree methods for finding active objects," *Advances in Space Research*, pp. 1955-1959, vol. 41, 2008.

[16] S. Kalmegh, "Analysis of WEKA data mining algorithm REPTree, simple cart and randomtree for classification of Indian news," *International Journal of Innovative Science, Engineering & Technology*, pp. 438-446, vol. 2, 2015.

Yu-Chen Chen was born in Taoyuan city, Taiwan (ROC) in 1992. She finished her bachelor's degree in industrial and systems engineering at Chung Yuan Christian University, Taoyuan city, Taiwan in 2014. She is a master student currently attending a double master degree program between the Department of Industrial Administration at Tokyo University of Science, Noda City, Japan and the Department of Industrial Engineering and Management at National Chiao Tung University, Hsinchu city, Taiwan. Her major fields of study at Tokyo University of Science are information engineering and machine learning.

Takashi Suzuki graduated from the Department of Mechanical Engineering, Tokyo University of Science, Katsushika-ku, Japan, 2012. Then he received Master of Engineering from the Graduate School of Engineering, Tokyo University of Science, 2014. He is a Doctoral course student in the Graduate School of Mechanical Engineering, Tokyo University of Science and a member of the collaborative research team between the Graduate School of Mechanical Engineering, Tokyo University of Science and the Department of Neurosurgery, Jikei University School of Medicine.

Masaaki Suzuki graduated from the Department of Physics, Gakushuin University, Toshima-ku, Japan, 2001. He then graduated from The University of Tokyo Graduate School, School of Engineering, Department of Quantum Engineering and Systems Science Doctoral course Completed program with degree, Bunkyo-ku, Japan, 2006. He was an assistant professor in Research into Artifacts, Center for Engineering (The University of Tokyo) from 2006 to 2010 and assistant professor in the Department of Nuclear Engineering and Management (The University of Tokyo) from 2010 to 2016. He has been an assistant professor at the Department of Industrial Administration, Faculty of Science and Engineering, Tokyo University of Science from 2016. His research interests are in the fields of computational engineering.

Hiroyuki Takao received his master's degree from the Department of Neurosurgery, Jikei University School of Medicine (JUSM), Minato-ku, Japan, 2003. He graduated from JUSM's neurosurgery doctoral course in 2010. He was a researcher (JUSM) from 2003 to 2007, an assistant professor (JUSM) from 2007 to 2012, an assistant researcher (University of California) from 2012 to 2013, and an associate professor (JUSM) for one month in 2013. He has been an associate professor at the Department of Neurosurgery and the Department of Innovation for Medical Informaion Techology JUSM since 2015. His recent research interests are developing IoT applications in ICT medicine research and creating software for analyzing cerebral aneurysm simulations.

Yuichi Murayama is a professor and chairman of Department of Neurosurgery at Jikei University School of Medicine, Tokyo, Japan. He was also a Professor of the Division of Interventional Neuroradiology, at UCLA School of Medicine. His academic interest is Stroke related innovation, including endovascular devices, such as Matrix coils or Onyx liquid embolic system, IT stroke network system, and Hybrid OR system. Currently he serves as a World Federation of Neurosurgical societies (WFNS) advisory panel of cerebrovascular diseases and therapy committee. He trains hybrid Neurosurgeons who can perform both open vascular technique and endovascular treatment.

Hayato Ohwada graduated from the Department of Industrial Administration, Faculty of Science and Technology, Tokyo University of Science, Noda City, Japan, in 1983. He also completed a Doctoral course from the Tokyo University of Science Graduate School, Division of Science and Engineering Industrial Administration, in Noda City, Japan, graduating in 1988. He was a research associate (Tokyo University of Science) from 1988 to 1998, a lecturer (Tokyo University of Science) from 1999 to 2000, and an associate professor (Tokyo University of Science) from 2001 to 2004. He has been a professor in the Department of Industrial Administration, Faculty of Science and Engineering, Tokyo University of Science since 2005. His research interests are in the fields of inductive logic programming and bioinformatics.