Blood Donor Classification Using Neural Network and Decision Tree Techniques

Wijai Boonyanusith and Phongchai Jittamai, *Member, IAENG*

***Abstract*—Blood donation is an essential activity to acquire blood as a raw material into the blood supply chain. It must be managed effectively together with other processes in blood management. In this research, the pattern of blood donors' behaviors based on factors influencing blood donation decision is conducted using online questionnaire. These factors, i.e., altruistic values, knowledge in blood donation, perceived risks, attitudes towards blood donation, and intention to donate blood, are analyzed to find out the possibilities for individuals to become blood donors. The surveyed data are used for machine learning techniques of Artificial Intelligence to classify the blood donor group into donors and non-donors. Moreover, the accuracy testing of the surveyed information is conducted using the Artificial Neural Network (ANN) and Decision Tree techniques in order to predict from a series of individual blood behavior data whether or not each individual is a donor. The results indicate that the accuracy, precision, and recall values of ANN technique are higher than those of the Decision Tree technique. Thus, the development of blood donor classification model leads to an improvement of the blood acquisition in Thailand.**

***Index Terms*—Classification, Blood Donor, Artificial Neural Network, Decision Tree**

I. Introduction

IN Thailand, the National Blood Center of the Thai Red Cross Society is responsible for blood and blood components management. Its main activities are acquiring blood supply from donors, blood screening and processing, blood storage and blood distribution. The National Blood Center performs blood collection from donors and distributed blood to various hospitals nationwide via Regional Blood Centers. Blood is distributed to hospitals according to their demands. Each Regional Blood Center is responsible to process blood and to prepare blood components such as plasmas and platelets for distribution to hospitals in its responsible network accurately and appropriately. Distribution is decided based on the level of available blood in the inventory and specific blood demands from patients in each hospital [1].

Blood collection from voluntary donors is a primary activity to acquire blood as raw material into blood supply chain. Generally, Blood Centers have a standard procedure to screen blood effectively by incorporating collection costs, and amounts of time and units, to obtain blood properly into consideration. The goal is to obtain blood that is safe for use in the subsequent activities. However, the main problem in blood collection is an inability to obtain sufficient blood to meet the patients' needs or a difficulty to balance blood demand and supply in the blood supply chain.

Problems in blood donation are mainly about insufficiency to collect blood to meet uncertain demands. In general, there is no plan to manage potential blood donors. With a good preparation, by classifying potential donors in such a way that intentions of donors to donate blood in the future can be determined, would greatly facilitate the blood acquisition to perform better. Response to blood requisition can be conducted efficiently and effectively, leading to life-saving of patients and reduction of relevant operating costs. Moreover, it will be beneficial for managing blood requisition on emergency requests. Hence, information from blood donor classification would suggest the behavior of blood donors whether or not they will donate again in the future. This information would greatly facilitate and improve blood donors recruiting process of the Blood Centers in Thailand.

Most research in the field uses only interview technique to classify donors. However, this research aims to use factors influencing potential donors to donate blood by incorporating the machine learning techniques using Artificial Neural Network (ANN) and Decision Tree to classify the surveyed data under the test by trusted external judgments.

II. Literature Review

A. Influenced Factors to Intention of Blood Donation

The study of the factors that influence the behavior of blood donors has been conducted extensively due to the significant impact of blood shortages to the survival of patients. Mostafa [2] introduced five factors leading to the study of blood donor behavior. These factors are (1) altruistic values, (2) perceived risks of blood donation, (3) blood donation knowledge, (4) attitudes toward blood donation, and (5) intention to donate blood.

In Thailand, there was a study in various reasons of factors that stimulate the altruistic value in donating blood [3]. People who have positive attitude are usually considerate of mankind and altruistic [4][5]. Perception of donors about risks in blood donating is one of the main reasons affecting a decision-making process in blood

Manuscript received July 16, 2012; Revised August 07, 2012. This work was supported in part by Suranaree University of Technology.

W. Boonyanusith is a Ph.D. student in the School of Information Technology, Suranaree University of Technology, Nakhon Ratchasima 30000, THAILAND (e-mail: boonyanusith@gmail.com).

P. Jittamai, Ph.D., is currently an Assistant Professor and a Chair of the School of Industrial Engineering, Suranaree University of Technology, Nakhon Ratchasima 30000, THAILAND (e-mail: jittamai@sut.ac.th).

donation [6] because it influences donor to have negative feeling toward blood donation [7]. Moreover, the fear of infection directly affects donors' intention to donate blood [8] presuming that those who donate blood may likely be infected more than those who do not [9]. On the other hand, those who donated blood in the past tend to have positive thought about blood donation [10] and may donate more in the future because they understand processes of blood donors screening better than the first-time donors [11]. Blood donors usually have much better positive attitude and good intention toward blood donation than those who have not donated blood before [12]. The intention to donate blood is one of the most important factors that can be used to predict behaviors in donating blood [13][14]. These aforementioned five factors are used in developing questionnaire to conduct individuals' opinions. This information is used for blood donor classification analysis in Thailand.

B. Classification Technique

The technique in discovering new patterns of large data sets is data mining. It can be used to extract knowledge from an existing data set and transform into a human-understandable structure for further use. It utilizes methods at the intersection of statistics, database systems, machine learning, and artificial intelligence [15]. ANN is a technique of data mining that is used to predict or classify data in the domain of ideas or feelings and behaviors of consumers effectively [16]. Its models in learning patterns of the data [17] to solve the problem of classification and clustering [18] are effective to analyze the marketing databases [19]. Multi-layer Perceptron is a popular and useful feedforward ANN model, which can be used to analyze data to classify the targeted group [20]. Moreover, Decision Tree is one of the useful techniques in classification by learning patterns of the data. It can display result graphically as a tree model in order to indicate each steps of decision process from input to output [21]. It uses CART [22] and C4.5 [23] algorithms for Tree-Building and Tree-Pruning in the Decision Tree process making.

Thus, this work aims to develop blood donor classification model from these aforementioned factors that influencing behaviors in blood donation and compare results in blood donor classification between ANN and Decision Tree techniques under the test by trusted external judgments.

III. Methodology

A. Online Questionnaire

At the first step of this work, questionnaire was used to collect data about feelings or opinions of individuals who become blood donors. Online questionnaire has been used to conduct data survey from 400 samples of people who are studying in 4 universities: Suranaree University of Technology, Nakhon Ratchasima Rajabhat University, Rajamangala University of Technology Isan, and Vongchavalitkul University. These universities are in Nakhon Ratchasima province, which is the largest province in the northeastern part of Thailand.

The questionnaire consists of two parts: the first part is about personal information including sex, age, level of education, weight and height, contact address, phone numbers, and the other part is 16 questions about their feelings or opinions why they decide to become blood donors. The scope of questionnaire was derived from the study of the factors influencing blood adopted from the work in the aforementioned literature review. The measurement format in this questionnaire follows the Likert-type scale, which is commonly used with queries individual attitude or feelings. This format is the opinion of respondents with 1 to 5 (strongly disagree, disagree, not sure, agree, and strongly agree). It also includes questions on both positive and negative combinations.

The factors used in this study aiming to analyze the intensity of the feelings or opinions of the individuals comprise of the following five factors:

- Altruistic values (ALT): There are four questions asking for attitudes about blood shortage reduction, benefit to the society, social ethics, and social belief.
- Knowledge in blood donation (KNL): There are four questions related to the knowledge asking for individual's age and body mass index (BMI) in blood donation along with the understanding screening and recruiting first-time blood donors.
- Perceived risks (RSK): There are three questions related to an individual's perception of standard procedure and safety control in blood donation along with the risk issues in donating blood.
- Attitudes in blood donation (ATT): There are three questions about attitudes in the satisfaction and the aspect of blood donation.
- Intention to donate blood (INT): There are two questions asking for the probability and intention to donate blood again in the future.

The survey was conducted from November, 2011 to March, 2012. The preliminary statistics were analyzed by calculating the median, standard deviation, and reliability testing (Cronbach' Alpha) of each factor. This statistical data is shown in Table I.

TABLE I
PRELIMINARY STATISTICS AND RELIABILITY OF THE QUESTIONNAIRE

Construct	Mean	SD	Cronbach's Alpha
ALT			
ALT 1	4.66	0.58	
ALT 2	4.75	0.46	**0.665**
ALT 3	3.71	1.07	
ALT 4	3.70	1.22	
KNL	-	-	-
RSK			
RSK 1	4.67	0.59	
RSK 2	4.45	0.89	**0.551**
RSK 3	4.15	1.06	
ATT			
ATT 1	4.71	0.51	
ATT 2	4.20	1.02	**0.669**
ATT 3	4.25	1.07	
INT			
INT 1	4.41	0.77	**0.597**
INT 2	4.20	0.95	
Total			**0.651**

ISBN: 978-988-19251-6-9
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

Statistics in Table I indicate that the altruistic values, perception of risks, attitudes in blood donation, and intention to donate blood have reliability values of Cronbach's Alpha of 0.665, 0.551, 0.669, and 0.597, respectively. There is no reliability test of knowledge in blood donation because all questions use "yes" or "no" answers. Finally, the reliability value of overall questionnaire equals to 0.651 of Cronbach's alpha which is in acceptable level (α>0.65). Thus, this questionnaire, which developed from such factors that influence the blood donation behaviors, can be used to explore behaviors of individuals at an appropriate level.

B. Classification Using Neural Network

To classify blood donor groups, it needs to prepare input data to the learning process by adjusted data obtained from questionnaires and made them as input attributes, according to the technique of ANN, in order to classify output as blood donor groups. The input data are shown in Table II.

TABLE II
EXAMPLES OF INPUT ATTRIBUTES AND DATA USED IN LEARNING PROCESS BY NEURAL NETWORK TECHNIQUE

ALT 1	ALT 2	ALT 3	ALT 4	KNL 1	KNL 2	KNL 3	KNL 4	RSK 1	RSK 2	RSK 3	ATT 1	ATT 2	ATT 3	INT 1	INT 2	Class
5	5	4	4	yes	yes	no	no	5	4	4	4	5	5	4	1	no
5	5	3	2	yes	yes	no	yes	5	4	4	5	5	5	4	2	yes
4	5	3	5	yes	yes	no	yes	4	1	4	5	5	5	1	2	no
5	5	5	5	yes	no	no	yes	5	4	4	5	5	5	5	5	yes
4	5	1	1	yes	yes	yes	yes	5	4	4	5	5	5	4	5	yes

Overall data includes 400 records using the five factors in learning process by ANN technique. This study focuses only on information of feelings or opinions that influence on blood donation behaviors only. However, basic personal information can be used for tracking blood donors for future blood donating. It is also used to develop the donor database system for better donor management.

IV. RESULTS

The section is divided into two parts. The first part aims to evaluate the model using ANN classification technique. The information retrieval criteria were used to analyze accuracy, precision, and recall values of the model, which can be expressed as (1), (2), and (3), respectively:

$$\text{Accuracy} = (TP + TN) / (TP + FP + FN + TN) \quad (1)$$
$$\text{Precision} = TP / (TP + FP) \quad (2)$$
$$\text{Recall} = TP / (TP + FN) \quad (3)$$

where
TP = donors in group D_j and model classified into group D_j
FP = donors not in group D_j and model classified into group D_j
FN = donors in group D_j and model not classified into group D_j
TN = donors not in group D_j and model not classified into group D_j
D_j = blood donor groups (donors and non-donors)

The other part compares the accuracy, precision, and recall values of the model using two techniques, ANN and Decision Tree.

A. Model Evaluation

The blood donor classification model was evaluated using ANN technique. It applied Multi-layer Perceptron and backpropagation algorithm [24] with the learning rate of 0.1 and the momentum was set equal to 0.1 in the learning process of input data with WEKA [25].

Input data consisting of 400 records were tested and used to construct network by 10 fold cross-validation. Its network architecture has 16 input nodes and 9 hidden layers to categorize the blood donors into two groups, donors (yes) and non-donors (no). The result of the classification is shown as a confusion matrix in Table III.

TABLE III
CONFUSION MATRIX IN CLASSIFICATION OF MODEL

Group	Donors	Non-Donors	Total
Donors	263	36	299
Non-Donors	59	42	101

Table III displays that the model is able to classify the output of blood donors from a total of 400 input records used in this experiment. Out of 299 records of confirmed donors, the model is able to classify 263 records correctly as confirmed donors: out of 101 records of non-donors, the model is able to classify 42 records correctly as non-donors.

Moreover, the classification of model was tested using precision and recall values and the results are shown in Table IV.

TABLE IV
PRECISION AND RECALL VALUES OF MODEL EVALUATION USING NEURAL NETWORK TECHNIQUE

Class	Precision	Recall
Donors	81.7 %	88.0 %
Non-Donors	53.8 %	41.6 %

Table IV shows that ANN model is able to classify input records as a blood donor group with the precision value of 81.7% and the recall value of 88.0%. It also has the precision and the recall values of 53.8% and 41.6%, respectively to classify them as a non-donor group. Accuracy value of the model equals to 76.25%. It can be claimed that the accuracy, precision, and recall values are high (over 75%) to classify blood donors appropriately.

This study also applied the Decision Tree technique to classify the same set of data into blood donor groups. Results from two techniques are compared and discussed.

B. Comparative Evaluation between Artificial Neural Network and Decision Tree

The model evaluation with Decision Tree technique in classification implemented J48 of the C4.5 algorithm in order to learn a set of input data using WEKA. The results of the model in classification are shown in Table V.

TABLE V
PRECISION AND RECALL VALUES OF MODEL EVALUATION USING DECISION TREE TECHNIQUE

Class	Precision	Recall
Donors	81.16 %	87.3 %
Non-Donors	52.5 %	41.6 %

ISBN: 978-988-19251-6-9
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

Table V shows that the Decision Tree model is able to classify input records as a blood donor group with the precision value of 81.16% and the recall value of 87.3%. It also has the precision and the recall values of 52.5% and 41.6%, to classify them as a non-donor group. Accuracy value of the model equals to 75.75%. Comparative evaluation of both techniques, using ANN and Decision Tree, is shown in Fig. 1.

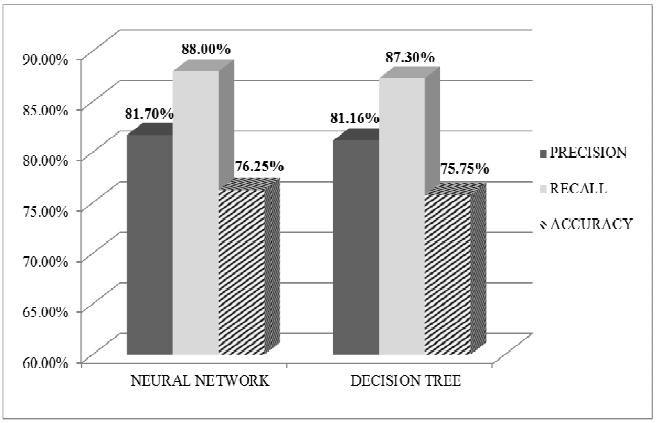

Figure 1. Precision, recall, and accuracy values of Neural Network and Decision Tree models in blood donor classification

It can be found that accuracy, precision, and recall values to classify blood donors using the ANN model are better than those values of the Decision Tree model. Thus, it can be stated that the ANN technique is a better approach to classify blood donors than the Decision Tree technique.

V. Conclusion

Blood donation is an essential activity to import a raw material into the blood supply chain. It must be managed effectively together with other processes in the blood supply chain management. The publicity of blood donation without the clear direction or the right target group of donors may cause adverse effects on higher cost of donation, time lost, and poor quality of blood. Applying information of consumer survey such as behaviors, feelings and opinions of the donors in blood donation can enhance the analysis of the feasibility of blood donation of each individual.

Five factors influencing blood donor behaviors suggested by Mostafa [2] are used as a framework to construct a model to classify donors. Those five factors include altruistic values, knowledge in blood donation, perceived risks, attitudes towards blood donation, and intention to donate blood. Online questionnaires were used in the survey among students in four universities in Nakhon Ratchasima province. A sample size of 400 responses has yielded a reliability scale in an acceptable significant level (greater than 0.65 of Cronbach's Alpha).

The accuracy test of donor group classification was done using the ANN technique with Multi-layer Perceptron function and backpropagation algorithm in order to predict the answer from a series of donors. As a result, it has been found that the model is able to classify donors into the blood donor group with the precision and recall values of 81.7% and 88%, respectively, and classify donors into the non-donor group with the precision and recall values of 53.8% and 41.6% respectively. The accuracy of this model to classify blood donor group is 76.25%.

Decision Tree technique is also used to compare the model with the ANN in order to analyze the accuracy of the classification of the model. The result has shown that the ANN model has higher accuracy value to classify blood donor groups than that of the Decision Tree model. It can be concluded that the ANN technique is a reliable method to learn data series of the five factors influencing the individual behavior of blood donation used for blood donor classification with high level of accuracy value.

In conclusion, Thailand still faces a challenge in acquiring enough blood to meet the inconsistent demands from the patients. The Thai Red Cross Society still needs to find a better approach to acquire blood and manage a group of donors. The blood donor classification model is a reliable method to help maintaining and managing a group of donors more effectively.

This research is a pilot study to incorporate the blood donor classification model to manage blood donors in Thailand. Machine learning techniques are used to classify a group of donors by testing with a data from potential donors in the northeastern province of the country. According to the preliminary statistical results of the factors affecting blood donor behavior, it can be claimed that reliability of the questionnaire and the sample is in an acceptable significant level. The data collected from donors can be used to develop the blood donor classification model appropriately. Although the sample size chosen in this study is only in the provincial level, the result shows that the ANN model has a relatively high accuracy value. This is an indication that the model is able to learn the pattern of blood donors from questionnaire with satisfactory results. Moreover, this study can be used as a prototype and expand the sample group in order to develop blood donor classification model in both the regional and the national levels, which will be beneficial for developing blood donor database system in different levels. This database system is advantageous for classifying potential blood donors and tracking them to donate blood again in the future. The obtained information can be used to determine potential blood donors and manage blood on the supply chain more effectively. Furthermore, this classification model and donor database system will contribute greatly for blood acquisition especially when there are emergency needs for blood for uses in the live-saving treatments.

VI. Future Works

Other factors affecting blood donation should be explored in order to better manage blood donors' behaviors. The work can be extended by enlarging the sample size of the study. Moreover, other machine learning techniques can be used comprehensively in analyzing the model and comparing the results.

Acknowledgment

We would like to thank the Regional Blood Center V of the Thai Red Cross Society (TRCS) and Suranaree University of Technology for their supports on this study.

REFERENCES

[1] The National Blood Center, Thai Red Cross Society.
[2] M. M. Mostafa, "Profiling blood donors in Egypt: a neural network analysis," *Expert Systems with Applications*, vol. 36, pp. 5031–5038, 2009.
[3] V. Wiwanikit, "A study on attitudes towards blood donation among people in a rural district, Thailand," *South Asian Journal of Tropical Medical Public Health*, vol. 31, pp. 609–611, 2000.
[4] J. Piliavin and P. Callero, *Giving blood: The development of an altruistic identity*. Baltimore: The Johns Hopkins University Press, 1991
[5] J. Piliavin, "Why do they give the gift of life? A review of research on blood donors since 1977," *Transfusion*, vol. 30, pp. 444–459, 1990.
[6] J. Allen and D. Butler, "Assessing the effects of donor knowledge and perceived risk on intention to donate blood," *Journal of Health Care Marketing*, vol. 13, pp. 26–33, 1993.
[7] D. Tscheulin and J. Lindenmeier, "The willingness to donate blood: an empirical analysis of socio-demographic and motivation-related determinants," *Health Services Management Research*, vol. 18, pp. 165–174, 2005.
[8] B. Robertson and D. McQueen, "Perceived risk of becoming infected with HIV by donating blood and changes in reported blood donation practice among the Scottish general public 1989–92," *AIDS Care*, vol. 6, pp. 435–442, 1994.
[9] J. Burnett, "Psychographic and demographic characteristics of blood donors," *Journal of Consumer Research*, vol. 8, pp. 62–66, 1981.
[10] G. Hosain, M. Anisuzzama, and A. Begum, "Knowledge and attitude towards voluntary blood donation among Dhaka University students in Bangladesh," *East African Medical Journal*, vol. 73, pp. 549–552, 1997.
[11] N. Zaller, *et al*., "Knowledge, attitude and practice survey regarding blood donation in a Northwestern Chinese city," *Transfusion Medicine*, vol. 15, pp. 277–286, 2005.
[12] S. Breckler and E. Wiggins, "Scales for the measurement of attitudes toward blood donation," *Transfusion*, vol. 29, pp. 401–409, 1989.
[13] H. Charng, J. Piliavin, and P. Callero, "Role identity and research action in the prediction of repeated behavior," *Social Psychology Quarterly*, vol. 51, pp. 303–317, 1988.
[14] M. Giles, *et al*., "An application of the theory of planned behavior to blood donation: The importance of self-efficacy," *Health Education Research*, vol. 19, pp. 380–391, 2004.
[15] K. Smith and J. Gupta, "Neural networks in business: Techniques and applications for the operations researcher," *Computers and Operations Research*, vol. 27, pp. 1023–1044, 2000.
[16] S. McKechnie, "Integrating intelligent systems into marketing to support market segmentation decisions," *Intelligent Systems in Accounting, Finance and Management*, vol. 14, pp. 117–127, 2006.
[17] C. Bishop, *Neural networks for pattern recognition*. New York: Oxford University Press, 1999.
[18] R. Sharda, "Neural networks for the MS/OR analyst: An application bibliography," *Interfaces*, vol. 24, pp. 116–130, 1994.
[19] J. Zahavi and N. Levin, "Applying neural computing to target marketing," *Journal of Direct Marketing*, vol. 11, pp. 5–22, 1997.
[20] T. Reutterer and B. Natter, "Segmentation-based competitive analysis with MULTICLUS and topology representing networks," *Computers and Operations Research*, vol. 27, pp. 1227–1247, 2000.
[21] J. R. Quinlan, "Induction of Decision Trees," *Machine Learning*, vol. 1, no. 1, pp. 81-106, 1986.
[22] L. Breiman, *et al.*, *Classification and Regression Trees.* CRC Press, New York, 1999.
[23] J. R. Quinlan, *C4.5: programs for machine learning*. Morgan Kaufmann, San Mateo, CA, 1993.
[24] C. Lim and T. Kirikoshi, "Predicting the effects of physician-directed promotion on prescription yield and sales uptake using neural networks," *Journal of Targeting, Measurement and Analysis for Marketing*, vol. 13, pp. 158–167, 2005.
[25] WEKA, University of Waikato, New Zealand. Available: http://www.cs.waikato.ac.nz/ml/weka/

ISBN: 978-988-19251-6-9
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)