EXCLI Journal 2013;12:885-893 – ISSN 1611-2156
Received: September 29, 2013, accepted: October 10, 2013, published: October 21, 2013

Original article:

MACHINE LEARNING APPROACHES FOR DISCERNING INTERCORRELATION OF HEMATOLOGICAL PARAMETERS AND GLUCOSE LEVEL FOR IDENTIFICATION OF DIABETES MELLITUS

Apilak Worachartcheewan[1], Chanin Nantasenamat[1,2]*, Pisit Prasertsrithong[2], Jakraphob Amranan[2], Teerawat Monnor[1], Tassaneya Chaisatit[3], Wilairat Nuchpramool[3], Virapong Prachayasittikul[2]*

[1] Center of Data Mining and Biomedical Informatics, Faculty of Medical Technology, Mahidol University, Bangkok 10700, Thailand
[2] Department of Clinical Microbiology and Applied Technology, Faculty of Medical Technology, Mahidol University, Bangkok 10700, Thailand
[3] International Center for Medical and Radiological Technology, Golden Jubilee Medical Center, Faculty of Medical Technology, Mahidol University, Bangkok 10700, Thailand

* Corresponding authors:
E-mail: chanin.nan@mahidol.ac.th; phone: +66 2 441 4371 ext. 2720, Fax: +66 2 441 4380
E-mail: virapong.pra@mahidol.ac.th; phone: +66 2 441 4376, Fax: +66 2 441 4380

ABSTRACT

Background: The aim of this study is to explore the relationship between hematological parameters and glycemic status in the establishment of quantitative population-health relationship (QPHR) model for identifying individuals with or without diabetes mellitus (DM).
Methods: A cross-sectional investigation of 190 participants residing in Nakhon Pathom, Thailand in January-March, 2013 was used in this study. Individuals were classified into 3 groups based on their blood glucose levels (normal, Pre-DM and DM). Hematological (white blood cell (WBC), red blood cell (RBC), hemoglobin (Hb) and hematocrite (Hct)) and glucose parameters were used as input variables while the glycemic status was used as output variable. Support vector machine (SVM) and artificial neural network (ANN) are machine learning approaches that were employed for identifying the glycemic status while association analysis (AA) was utilized in discovery of health parameters that frequently occur together.
Results: Relationship amongst hematological parameters and glucose level indicated that the glycemic status (normal, Pre-DM and DM) was well correlated with WBC, RBC, Hb and Hct. SVM and ANN achieved accuracy of more than 98 % in classifying the glycemic status. Furthermore, AA analysis provided association rules for defining individuals with or without DM. Interestingly, rules for the Pre-DM group are associated with high levels of WBC, RBC, Hb and Hct.
Conclusion This study presents the utilization of machine learning approaches for identification of DM status as well as in the discovery of frequently occurring parameters. Such predictive models provided high classification accuracy as well as pertinent rules in defining DM.

Keywords: Diabetes mellitus, glucose, hematologic parameters, quantitative population-health relationship, QPHR, data mining

EXCLI Journal 2013;12:885-893 – ISSN 1611-2156
Received: September 29, 2013, accepted: October 10, 2013, published: October 21, 2013

INTRODUCTION

Diabetes mellitus (DM) is defined as a group of metabolic disorders characterized by increasing blood glucose or hyperglycemia that occur as a result of defects in insulin secretion, insulin action or insulin resistance (IR) (Salsali and Nathan, 2006). DM can be classified as type 1 (i.e. non-secretion of insulin) and type 2 (i.e., defined by IR) (Salsali and Nathan, 2006). DM is a complicated disease affecting multiple tissues in the body as it impairs organ function or biochemical parameters leading to expression of clinical signs and symptoms (Salsali and Nathan, 2006). Finally, without proper diagnosis or treatment, morbidity or mortality can occur.

Hematological parameters, such as white blood cell (WBC), red blood cell (RBC), hemoglobin (Hb) and hematocrit (Hct) that had previously been found to be correlated with IR and metabolic syndrome (MS) (Chen et al., 2006; Choi et al., 2003; Jung et al., 2013; Kawamoto et al., 2013; Wang et al., 2004), a condition predisposing individuals to the development of DM and/or cardiovascular diseases (Alberti et al., 2009). These parameters are implied to be related to glycemic status in which there may be involvement with inflammatory response or blood flow in the body. Therefore, those parameters may be used for identifying individuals with or without glycemic condition.

Machine learning techniques are computational approaches that are used in extracting pertinent knowledge from large amounts of data and it has previously been applied for various clinical decision-making process such as metabolic syndrome (Kim et al., 2012a; Worachartcheewan et al., 2013), diabetes mellitus (Quentin-Trautvetter et al., 2002; Yu et al., 2010), cancer (Nahar et al., 2011) and hypertension (Shin et al., 2010). The aim of this study is to classify individuals as normal, Pre-DM and DM using the quantitative population-health relationship (QPHR) approach employing machine learning approaches such as support vector machine, artificial neural network and association rule analysis.

MATERIAL AND METHODS

Sample population

A sample population of 190 individuals (i.e., comprising of 71 males and 119 females) residing in Nakhon Pathom, Thailand and receiving health check-up from the International Center for Medical and Radiological Technology, Golden Jubilee Medical Center, Faculty of Medical Technology, Mahidol University during the duration of January-March, 2013 was used in this study.

Blood chemical measurements

Individuals were described by a set of health parameters encompassing blood chemical tests including blood glucose, total cholesterol (Chol), triglyceride (TG), low-density lipoprotein cholesterol (LDL-C), high-density lipoprotein cholesterol (HDL-C) while hematological parameters comprising of WBC, RBC, Hb and Hct. These parameters were determined by standard methods. Blood chemistry testings were stratified according to guidelines of the WHO (Wilson, 2009). Hematological parameters were divided into four groups based on their quartiles.

Machine learning analysis

Support vector machine (SVM) and artificial neural network (ANN) implementing the John Platt's Sequential Minimal Optimization algorithm (Witten et al., 2011) and the back-propagation algorithm (Nantasenamat et al., 2007), respectively, were employed in constructing classification models using WEKA, version 3.4.5 (Witten et al., 2011). In SVM models, the input data were transformed to higher dimensional space by means of the radial basis function kernel (Worachartcheewan et al., 2013) Hematological parameters and glucose level were used as input variables and the DM status (i.e., normal, Pre-DM and DM) was used as the output variable. Data sampling

EXCLI Journal 2013;12:885-893 – ISSN 1611-2156
Received: September 29, 2013, accepted: October 10, 2013, published: October 21, 2013

was performed by means of ten-fold cross-validation (10-fold CV).

Statistical analysis

Statistical analysis was performed using SPSS Statistics 18.0 (SPSS Inc. USA) in which *P*-value of < 0.05 was considered statistically significant. The predictive performance of the classification models was evaluated from its accuracy. In addition, association analysis (AA) was performed in SPSS Clementine, version 11.1 (SPSS Inc., USA) employing the *Apriori* algorithm (Agrawal et al., 1993) to discover frequently occurring parameters in individuals with DM. Association rules was obtained by using minimum support and confidence values of $min_{sup} = 5$ % and $min_{conf} = 70$ %, respectively.

RESULTS

Population characteristics and data preprocessing

Table 1 shows the biochemical features of 190 participants stratified by their glucose levels: normal for glucose < 100 mg/dL (i.e., comprising of 33 males and 74 females), Pre-DM for glucose 100 – 125 mg/dL (i.e., composed of 32 men and 27 women) and DM for glucose ≥ 126 mg/dL (i.e., comprising of 11 males and 13 females). In the DM group, the average values of glucose, TG and WBC were found to increase while Chol, HDL-C and LDL-C were found to decrease when compared to the other two groups. Interestingly, TG and HDL-C levels were shown to increase and decrease, respectively, in the Pre-DM and DM groups when compared to the normal group. Such observation coincides with the known fact that individuals with MS have increased TG and decreased HDL-C levels (Alberti et al., 2009). Furthermore, Hb, Hct and RBC were non-significant in the three groups. Box plots of biochemical parameters are presented in Figure 1. As shown in Figure 2, an intercorrelation matrix was constructed to discern relationships amongst investigated descriptors and it was found that the glycemic status was correlated with RBC, Hb and Hct as observed from the increasing correlation from the normal group to the Pre-DM group and finally the DM group.

Table 1: Clinical characteristics of sample population stratified by glycemic status

	Normal	Pre-diabetes	Diabetes	*P*-value
Case number	107 (56.32)	59 (31.05)	24 (12.63)	-
Male	33 (17.37)	27(16.84)	11 (5.79)	-
Female	74 (38.95)	32 (14.21)	13 (6.84)	-
Glucose (mg/dL)	91.82 ± 4.93	108.25 ± 6.02	185.29 ± 67.87	< 0.05
Cholesterol (mg/dL)	202.89 ± 43.42	188.90 ± 36.17	178.38 ± 40.75	< 0.05
Triglyceride (mg/dL)	112.89 ± 57.43	117.68 ± 46.29	144.54 ± 79.23	0.053
HDL-C (mg/dL)	59.93 ± 16.92	56.44 ± 15.87	51.17 ± 14.82	< 0.05
LDL-C (mg/dL)	125.96 ± 39.18	114.76 ± 28.66	105.00 ± 34.92	< 0.05
WBC ($\times 10^9$/L)	6.16 ± 1.74	6.69 ± 1.96	7.28 ± 2.14	< 0.05
RBC ($\times 10^{12}$/L)	4.56 ± 0.67	4.57 ± 0.63	4.49 ± 0.69	0.728
Hb (g/dL)	12.87 ± 1.39	12.87 ± 1.57	12.48 ± 1.99	0.179
Hct (%)	39.80 ± 4.44	39.80 ± 4.95	38.55 ± 6.09	0.151

Data was expressed as mean ± SD or as percentages. HDL-C: high-density lipoprotein cholesterol, LDL: Low-density lipoprotein cholesterol, WBC: white blood cell, RBC: red blood cell, Hb: hemoglobin and Hct: hematocrit.

EXCLI Journal 2013;12:885-893 – ISSN 1611-2156
Received: September 29, 2013, accepted: October 10, 2013, published: October 21, 2013

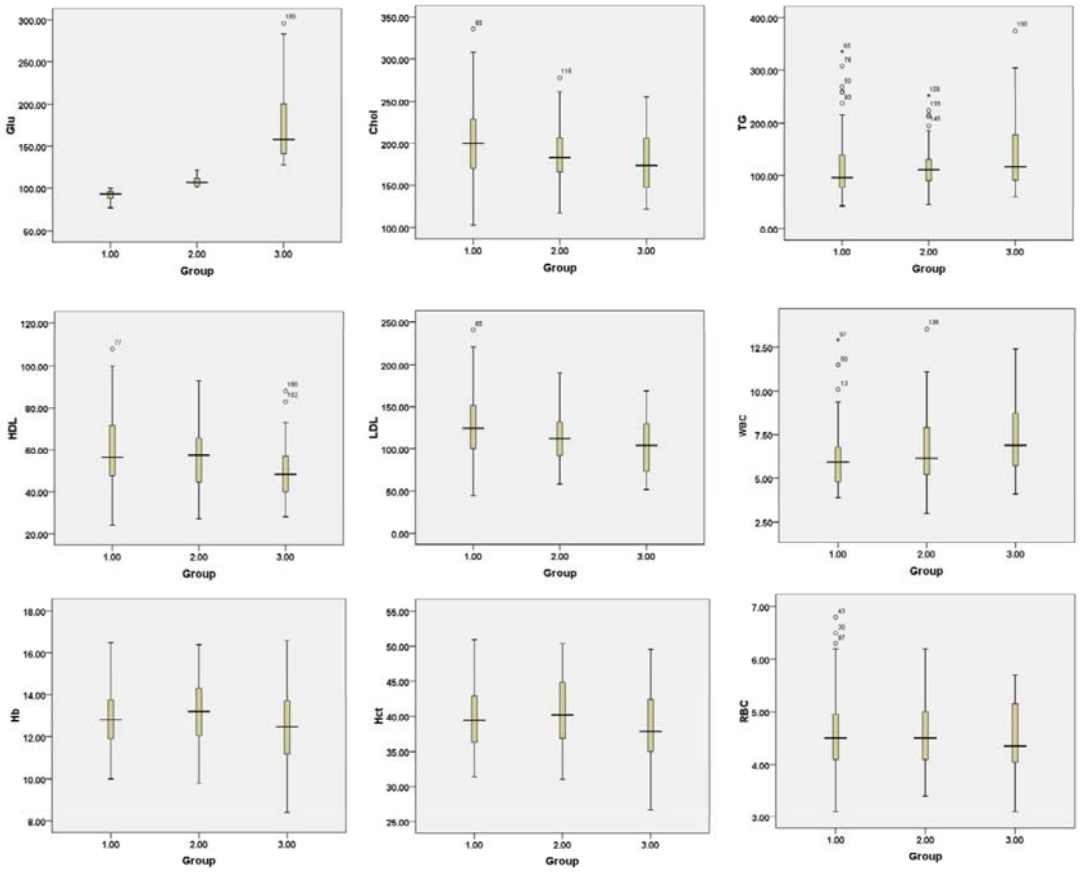

Figure 1: Box plots of biochemical parameters were stratified by their glycemic status to normal (a), Pre-DM (b) and DM (c) groups

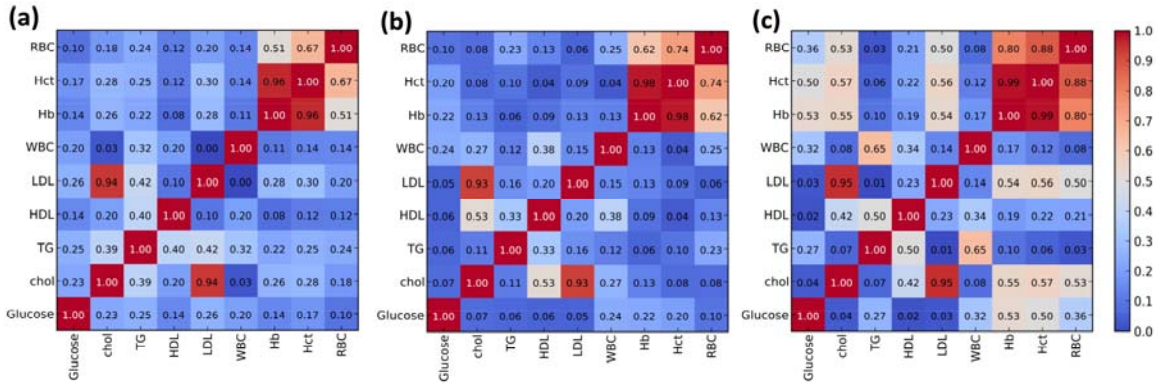

Figure 2: Intercorrelation matrix of biochemical and hematological parameters for normal (a), Pre-DM (b) and DM (c) groups

Classification via QPHR modeling

Hematological parameters comprising of WBC, RBC, Hb, Hct and glucose were used as input variables while DM, Pre-DM and normal groups (i.e., as classified by their glucose level; Table 1) were used as the dependent variable. These variables were used in the construction of classification models using SVM and ANN methods. Furthermore, WBC, RBC, Hb and Hct parameters were binned into four groups on the basis of quartiles while other blood

EXCLI Journal 2013;12:885-893 – ISSN 1611-2156
Received: September 29, 2013, accepted: October 10, 2013, published: October 21, 2013

chemistry parameters were stratified according to the WHO guidelines (Table 2). These parameters were used in the discovery of biochemical parameters that frequently occur together with or without DM via the use of AA via the *Apriori* algorithm.

Table 2: Binning of biochemical parameters. Each parameter was stratified according to guidelines of the WHO.

Factor		Male (n=71)	*n*	Female (n=119)	*n*
Glucose	glu_1	< 100	33 (46.48)	< 100	74 (62.19)
	glu_2	100-125	27 (38.03)	100-125	32 (26.89)
	glu_3	≥ 126	11 (15.49)	≥ 126	13 (10.92)
Cholesterol	chol_1	< 200	46 (64.79)	< 200	60 (50.42)
	chol_2	200-239	13 (18.31)	200-239	41 (34.45)
	chol_3	≥ 240	12 (16.90)	≥ 240	18 (15.13)
Triglyceride	TG_1	< 150	54 (76.06)	< 150	93 (78.15)
	TG_2	150-199	10 (14.08)	150-199	16 (13.45)
	TG_3	200-299	5 (7.04)	200-299	8 (6.72)
	TG_4	≥ 300	2 (2.82)	≥ 300	2 (1.68)
LDL-C	LDL_1	< 100	25 (35.21)	< 100	33 (27.73)
	LDL_2	100-129	23 (32.39)	100-129	41 (34.45)
	LDL_3	130-159	14 (19.72)	130-159	28 (23.53)
	LDL_4	≥ 160	10 (14.08)	≥ 160	17 (14.29)
HDL-C	HDL_1	≤ 40	13 (18.31)	≤ 50	38 (31.93)
	HDL_2	> 40	58 (81.69)	> 50	81 (68.07)
WBC	WBC_1	< 5.5	18 (25.35)	< 4.8	34 (28.57)
	WBC_2	5.6-6.4	20 (28.17)	4.9-5.9	26 (21.85)
	WBC_3	6.5-8.2	18 (25.35)	6.0-7.2	33 (27.73)
	WBC_4	> 8.2	15 (21.13)	> 7.2	26 (21.85)
RBC	RBC_1	< 4.3	18 (25.35)	< 4.0	31 (26.05)
	RBC_2	4.4-5.0	25 (35.21)	4.1-4.4	37 (31.09)
	RBC_3	5.1-5.3	15 (21.13)	4.5-4.7	25 (21.01)
	RBC_4	> 5.3	13 (18.31)	> 4.7	26 (21.85)
Hb	Hb_1	< 12.6	18 (25.35)	< 11.6	31 (26.05)
	Hb_2	12.7-14.2	20 (28.17)	11.7-12.4	31 (26.05)
	Hb_3	14.3-15.0	16 (22.54)	12.5-13.2	28 (23.53)
	Hb_4	> 15	17 (23.94)	> 13.2	29 (24.37)
Hct	Hct_1	< 38.8	18 (25.35)	< 36.1	30 (25.21)
	Hct_2	38.9-44.2	18 (25.35)	36.2-38.0	30 (25.21)
	Hct_3	44.3-47.0	18 (25.35)	38.1-40.6	30 (25.21)
	Hct_4	> 47.0	17 (23.94)	> 40.6	29 (24.37)

HDL-C: high-density lipoprotein cholesterol, LDL: Low-density lipoprotein cholesterol, WBC: white blood cell, RBC: red blood cell, Hb: hemoglobin and Hct: hematocrit. Data in parentheses is expressed as percentages.

EXCLI Journal 2013;12:885-893 – ISSN 1611-2156
Received: September 29, 2013, accepted: October 10, 2013, published: October 21, 2013

Support vector machine

In development of SVM classification models, C and γ parameters were optimized by means of a two-step process in which a coarse global grid search was followed by a refined local grid search. The former searches values from 2^{-19} to 2^{19} using an incremental increase in step size of 2^2, which gave $C = 2^{19}$ and $\gamma = 2^{-9}$ as global grid parameters affording 97.37 % of accuracy. A subsequent local grid search investigating regions spanning 2^{17} to 2^{23} was performed for the C parameter while the range of 2^{-11} to 2^{-7} was performed for the γ parameter using step sizes of $2^{0.25}$. Results indicated that optimal values for C and γ parameters were 2^{23} and $2^{-8.5}$, respectively, providing accuracies of 100 % and 98.42 % for the training set and 10-fold CV set, respectively, as shown in the confusion matrix presented in Table 3.

Artificial neural network

In constructing ANN classification models, ANN parameters were optimized and it was found that the optimal values for the number of hidden nodes, learning epochs, learning rate and momentum are 6, 10000, 0.1 and 0.3, respectively. The resulting model afforded accuracies of 100 % and 98.42 % for the training set and 10-fold CV set, respectively, which shows similar prediction results as that of SVM (Table 3).

Association rule analysis

AA was employed to discover frequently occurring variables in individuals with DM or without DM. AA analysis gave rise to 533 rules that can be stratified as follows: 411 rules for the normal group, 111 rules for the Pre-DM group and 11 rules for the diabetes group as provided in supplementary information (Tables S1-S3).

It is observed that rules for the diabetes group involved only abnormal glucose level 3 corresponding to glucose level of ≥ 126 mg/dL. Interestingly, rules for Pre-DM group are associated with increasing WBC, RBC, Hb and Hct affording levels of 2, 3 and 4 and glucose level of 2.

DISCUSSION

IR is a condition in which target cells are not responsive to insulin levels in circulation and this leads to the development of diseases such as MS, chronic inflammation, diabetes and cardiovascular diseases (Salsali and Nathan, 2006; Alberti et al., 2009). In considering hematological parameters, immune cells such as WBC may be involved in inflammatory response in which the adipose tissue is a target in IR. Subsequently, the adipose tissue secretes inflammatory factors such as cytokines to activate

Table 3: Summary of predictive performance for diabetes mellitus identification using support vector machine

	Support vector machine			
	Normal	**Pre-DM**	**DM**	**Accuracy**
Training set				100
Normal	107	0	0	
Pre-DM	0	59	0	
DM	0	0	24	
10-fold CV				98.42
Normal	107	0	0	
Pre-DM	0	58	1	
DM	0	2	22	

the increase of WBC (Bermann and Sypniewska, 2013). Hct and Hb have been documented to increase the levels of parameters related to high blood viscosity, consequently leading to a decrease in blood flow (de Simone et al., 1990) and an increase in blood pressure (Kutlu et al., 2009) as found in DM and Pre-DM groups, respectively. Furthermore, RBC was also found to be correlated with glycemic condition (Chen et al., 2006; Choi et al., 2003; Jung et al., 2013; Kawamoto et al., 2013; Wang et al., 2004) in which the mechanism of increasing RBC indices in the presence of IR is not completely understood, however, it may be deduced to be involved in the increase of erythropoiesis in peripheral blood or involved in reducing blood flow and rising viscosity thereby leading to elevated RBC count (Kawamoto et al., 2013).

Furthermore, hematological parameters were found to be correlated with glycemic conditions (Figure 2), which coincides with previous findings (Chen et al., 2006; Choi et al., 2003; Jung et al., 2013; Kawamoto et al., 2013; Wang et al., 2004). Particularly, RBC, Hb and Hct were shown to exhibit significant association with glycemic status (i.e., normal, Pre-DM and DM) as shown in Figure 2. Moreover, WBC was found to increase in both Pre-DM and DM groups (Table 1). Therefore, hematological parameters were used to classify individuals as having or not having DM.

Herein, the QPHR approach had successfully been shown to afford robust classification of glycemic status (i.e., normal, Pre-DM and DM) as a function of health parameters. The approach enables the correlation of biomedical parameters with their respective DM status. QPHR is a data mining approach previously termed by us and had been successfully employed in classifying MS (Worachartcheewan et al., 2010, 2013) while relevant effort had been shown to be useful in classifying DM (Quentin-Trautvetter et al., 2002; Yu et al., 2010). Previously, these methods have been shown by us to yield accuracies of 91 – 98 % in the classification of MS status (Worachartcheewan et al., 2013). Interestingly, AA identified abnormalities in hematological parameters (i.e., WBC, RBC, Hb and Hct) in the Pre-DM group while abnormal level of glucose (i.e. glucose level 3) was found in the DM group. Previously, AA was used to analyze the comorbidity in patients with type 2 DM (Kim et al., 2012b) and MS (Worachartcheewan et al., 2013) as to understand the correlation between biomedical parameters with diseases. Considering that the level of glucose was already used for labeling individuals as having DM or non-DM, it was therefore pertinent in the resulting DM identification. The inclusion of hematological parameters (i.e., WBC, RBC, Hb and Hct) in classifying DM status still led to an accuracy of more than 98 %. Therefore, it may be implied that hematological parameters are important variables together with glucose level for the identification of DM status.

The QPHR study performed herein for the first time presents the utilization of complete blood cell count parameters (i.e., WBC, RBC, Hb and Hct) as descriptors in classification of DM status. Results support the fact that hematological parameters and glycemic condition are correlated, particularly, the strongest correlation was observed for the DM status group as corroborated by previous studies (Chen et al., 2006; Choi et al., 2003; Jung et al., 2013; Kawamoto et al., 2013; Wang et al., 2004). Moreover, AA analysis suggests strong correlation between high levels of hematological parameters with the Pre-DM group (Table S2 in supplementary information). The machine learning approaches (i.e., SVM and ANN) employed in this study have been shown to be capable of correctly classifying the DM status affording accuracies of more than 98 %. In addition, rules obtained from AA analysis may be used as guidelines for the prevention of individuals at risk for the development of DM.

ACKNOWLEDGEMENTS

The annual budget grant of Mahidol University (B.E. 2556-2558) is gratefully acknowledged for supporting this research. The authors thank the International Center for Medical and Radiological Technology, Golden Jubilee Medical Center, Faculty of Medical Technology, Mahidol University for the data set used in this study.

REFERENCES

Agrawal R, Imilienski T, Swami A. Mining association rules between sets of items in large databases. Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data 1993:207-16.

Alberti KG, Eckel RH, Grundy SM, Zimmet PZ, Cleeman JI, Donato KA et al. Harmonizing the metabolic syndrome: a joint interim statement of the International Diabetes Federation Task Force on epidemiology and prevention; National Heart, Lung, and Blood Institute; American Heart Association; World Heart Federation; International Atherosclerosis Society; and International Association for the Study of Obesity. Circulation 2009;120:1640-5.

Bermann K, Sypniewska G. Diabetes as a complication of adipose tissue dysfunction. Is there a role for potential new biomarkers? Clin Chem Lab Med 2013;51: 177-85.

Chen L-K, Lin M-H, Chen Z-J, Hwang S-J, Chiou S-T. Association of insulin resistance and hematologic parameters: study of a middle-aged and elderly Chinese population in Taiwan. J Chin Med Assoc 2006;69:248-53.

Choi KM, Lee J, Kim YH, Kim KB, Kim DL, Kim S et al. Relation between insulin resistance and hematological parameters in elderly Koreans-Southwest Seoul (SWS) Study. Diabetes Res Clin Pract 2003;60: 205-12.

de Simone, Devereux RB, Chien S, Alderman MH, Atlas SA, Laragh JH. Relation of blood viscosity to demographic and physiologic variables and to cardiovascular risk factors in apparently normal adults. Circulation 1990;81:107-17.

Jung C-H, Lee W-Y, Kim B-Y, Park SE, Rhee E-J, Park C-Y et al. The risk of metabolic syndrome according to the white blood cell count in apparently healthy Korean adults. Yonsei Med J 2013;54:615-20.

Kawamoto R, Tabara Y, Kohara K, Miki T, Kusunoki T, Abe M, Katoh T. Hematological parameters are associated with metabolic syndrome in Japanese community-dwelling persons. Endocrine 2013;43:334-41.

Kim TN, Kim JM, Won JC, Park MS, Lee SK, Yoon SH et al. A decision tree-based approach for identifying urban-rural differences in metabolic syndrome risk factors in the adult Korean population. J Endocrinol Invest 2012a;35:847-52.

Kim HS, Shin AM, Kim MK, Kim YN. Comorbidity study on type 2 diabetes mellitus using data mining. Korean J Intern Med 2012b;27:197-202.

Kutlu M, Sonmez A, Genc H, Erdem G, Tapan S, Celebi G et al. Relationship between hemoglobin and CD40 ligand in prediabetes. Clin Invest Med 2009;32:E244.

Nahar J, Tickle KS, Ali AB, Chen YP. Significant cancer prevention factor extraction: an association rule discovery approach. J Med Syst 2011;35:353-67.

Nantasenamat C, Isarankura-Na-Ayudhya C, Tansila N, Naenna T, Prachayasittikul V. Prediction of GFP spectral properties using artificial neural network. J Comput Chem 2007;28:1275-89.

Quentin-Trautvetter J, Devos P, Duhamel A, Beuscart R. Assessing association rules and decision trees on analysis of diabetes data from the DiabCare program in France. Stud Health Technol Inform 2002;90:557-61.

Salsali A, Nathan M. A review of types 1 and 2 diabetes mellitus and their treatment with insulin. Am J Ther 2006;13:349-61.

Shin AM, Lee IH, Lee GH, Park HJ, Park HS, Yoon KII et al. Diagnostic analysis of patients with essential hypertension using association rule mining. Healthc Inform Res 2010;16:77-81.

Wang YY, Lin SY, Liu PH, Cheung BM, Lai WA. Association between hematological parameters and metabolic syndrome components in a Chinese population. J Diabetes Complications 2004;18:322-7.

Wilson DD. Manual of laboratory & diagnostic tests. New York: McGraw-Hill, 2009.

Witten IH, Frank E, Hall MA. Data mining: practical machine learning tools and techniques (2nd ed.). San Francisco, CA: Morgan Kaufmann, 2011.

Worachartcheewan A, Nantasenamat C, Isarankura-Na-Ayudhya C, Pidetcha P, Prachayasittikul V. Identification of metabolic syndrome using decision tree analysis. Diabetes Res Clin Pract 2010;90:e15-8.

Worachartcheewan A, Nantasenamat C, Isarankura-Na-ayudhya C, Prachayasittikul V. Quantitative population health relationship (QPHR) for assessing metabolic syndrome. EXCLI J 2013;12:569-83.

Yu W, Liu T, Valdez R, Gwinn M, Khoury M. Application of support vector machine modeling for prediction of common diseases: the case of diabetes and pre-diabetes. BMC Medical Informatics and Decision Making 2010;10:16-22.