Classification Method for Thai Elderly People Based on Controllability of Sugar Consumption

Punnarumol Temdee
School of Information Technology
Mae Fah Luang University
Chiang Rai, Thailand
Punnarumol@ mfu.ac.th

ChuanHui He
School of Information Technology
Mae Fah Luang University
Chiang Rai, Thailand
5931301051@lamduan.mfu.ac.th

Marzia Hoque Tania
Medical Technology Research Center
Anglia Ruskin University
Chelmsford, United Kingdom
marzia.hoque-tania1@anglia.ac.uk

***Abstract*— Nowadays, the number of Thai elders is rapidly increasing among world elderly population, how to keep their health is a major concern. Cardiovascular Diseases (CVDs) which are severe disease for Thai have higher mortality than cancers, and elderly people have a higher possibility to predispose CVDs. Hence, the risk factors of CVDs should be addressed. Obesity, as one of risk factors of CVDs, seriously affects Thai elders' wellbeing; excessive sugar consumption is a way leading to overweight and obesity. The amount of consumed sugar by Thai is much higher than the standard sugar consumption, and it also could cause many other diseases. Therefore, this paper proposes a classification method for elderly group who have the potential to control their blood sugar in order to prevent them from sugar overconsumption. This paper explored machine learning algorithms to find an appropriate classification method for elderly data. Artificial neuron network and K-nearest neighbors are applied for classifying elderly groups. Glycated hemoglobin (HbA1c) and fasting plasma glucose (FPG) are the noninvasive measurements of evaluating blood sugar, based on the two measurements, the 242 data from 121 elderly people are divided into two groups which are controllable group and uncontrollable group. The result indicates that artificial neuron network is more suitable for the dataset with 70.59% accuracy as compared to the accuracy of K-nearest neighbors.**

Keywords*— *Classification method, Thai senior, Cardiovascular Disease, Sugar consumption, Artificial Neuron Network, K-Nearest Neighbors

I. Introduction

Thailand, considered as a developing country with a potent economic growth, is now seriously facing the problem of an aging society. According to its demographic, the number of senior populations (aged 60 and older) dramatically increased from 1.2 million as 4.6% of Thai population in 1960 to 8 million as 11.9% in 2010 and is projected to reach 12 million within 2020 [1]. Their well-being thus becomes a major concern for preventing diseases such as Cardiovascular Diseases (CVDs), Respiratory Diseases, Blood Disorders, Renal, and Lower Urinary Tract Disorders [2].

CVDs are a list of disorders related to the heart and blood vessels which can lead to coronary heart disease, cerebrovascular disease, peripheral arterial disease, rheumatic heart disease, congenital heart disease, and deep vein thrombosis and pulmonary embolism. Importantly, CVDs are remarked as the leading cause of death worldwide with approximate 17.9 million deaths in 2016 [3]. In Thailand, CVDs accounted for 23% of mortality which is higher than cancers [4]. About estimated 54,000 people who died from CVDs accounted for 17% of the total deaths in 1998 [5]. Moreover, Thailand population with increasing age accompanies high probability of predisposing CVDs, which means the older Thai people become, the higher possibility of developing CVDs increases [6]. Therefore, it is important to focus on the risk factors of CVDs.

Many manageable risk factors of CVDs have been potentially addressed, for instance, blood pressure, blood cholesterol, smoking, obesity, and diabetes [5]. According to the World Health Organization (WHO), the projected obesity for Thailand is higher than the global target [4], and Thailand has the highest prevalence of overweight and obese adults among Asian countries [7]. Approximate one-third of elder Thais (age ≥ 60) are obese (BMI ≥ 25 kg/m^2) [8], older Thai have more frequent glucose and obesity than younger Thai [5], which considered as a significant risk factor of CVDs for Thai senior. Excessive sugar consumption results in superfluous discretionary calories which conduce to overweight and obesity [9].

Consuming in large quantities of sugar has increased yearly by Thai people, sugar consumption rose almost trebly from 12.7 kilograms per person in 1983 to 31.2 kilograms in 2009 [10]. 6 teaspoons of sugar per day are recommended for individuals while Thai people consume about 26 teaspoons per day that are more than quadruple than the suggested sugar intake [11]. Sugar overconsumption not only influences on increasing calories which could lead to overweight and obesity, but also rises blood sugar causing many diseases related to metabolic syndrome, which involves hypertension, diabetes mellitus, myocardial infarction, dyslipidemia, pancreatitis, and hepatic dysfunction [12]. Hence, excessive sugar intake is a serious concern for the Thai government at present. Group classification for elderly people based on their sugar consumption may be useful for their health.

Classification is beneficial to making decisions and widely used for solving decision problems embracing a specific object into a suitable group [13]. There are many classification methods nowadays, artificial neuron network (ANN) and K-nearest neighbors (KNN) are broadly used in plenty of scientific fields for classifying and predicting problems, for example, business, education, health, and physiological signals, etc. Using two of these classifiers has obtained great performance with high accuracy on each filed. Therefore, in this paper, ANN and KNN are used for classifying groups particularly for senior citizens who have a high potential of controlling blood sugar based on sugar consumption in order to avoid excessive sugar intake.

II. Literature Review

As mentioned previously, normal sugar consumption is acceptable for daily intake, but excessive sugar consumption can raise blood glucose level that causes many chronic diseases. In order to classify elderly groups who can possibly control their blood sugar based on sugar consumption, the measurements of blood sugar should be focused on. First is the measurement of glucose in the blood, plasma glucose is measured after filtration from the red blood cells; the amount of all blood glucose is called fasting plasma glucose (FPG). FPG is recommended for checking blood sugar level for diabetes by both WHO and American Diabetes Association (ADA), because FPG is easy to use and lower cost, and it also can substantially differentiate a group that possibly increases early mortality and risk of microvascular and CVDs [14][15]. Normal FPG should be less than 110 mg/dl (6.1 mmol/l), and the FPG boundary between diabetes and non-diabetes is 126mg/dl (7.0 mmol/l) [16]. Another measurement is glycated hemoglobin (HbA1c). Since blood sugar changes frequently within a day, HbA1c measures average blood glucose concentration during 2-3 months. Using HbA1c is convenient and quick for monitoring glycemic because patient's blood sample can be drawn any time and patients do not need any preparation such as fasting. Moreover, HbA1c is important in glycemic management for identifying people whose blood sugar is at higher risk for predisposing diabetes. Hence it was recommended to use for diagnosing blood sugar by ADA in 2009. The cutoff point of HbA1c is 6.5%, which means a person with HbA1c over 6.5% is deemed to a diabetes patient whose glycemic is quite high [14][16]. Moreover, FPG and HbA1c are non-invasive methods to measure blood sugar.

Machine learning-based grouping of elderly people can help health care professionals to convincingly recognize patients who can or cannot control their blood glucose based on sugar consumption. The classification methods should also be paid attention to as there are many classification methods. ANN and KNN are widely used in many domains, especially for physiological signals. ANN has been prevalently used in classifying pattern recognition, especially for bio-signals. For instance, Galvanic Skin Response (GSR) and Photoplethysmography (PPG) are used for analyzing stress pattern of automotive drivers in order to prevent road accidents, and ANN performs well with 94.92% specificity rate [17]. ANN classifies heart diseases that are pulmonary edema and respiratory failure with high accuracy by using seven physiological signals including Mean Arterial Blood Pressure (MABP), Systolic Blood Pressure (SBP), Diastolic Blood Pressure (DBP), Heart Rate (HR), Pulse Rate (Pulse), Respiration Rate (RESP), and Oxygen Saturation (SpO_2) [18]. Classifying cardiac abnormalities attains excellent performance by using Heart Rate Variability (HRV) signals as a non-invasive measurement for diagnosis [19]. ANN provides 100% accuracy for the classification of Electrocardiogram (ECG) signals to check cardiac health [20]. Recognizing hand gesture using Electromyography (EMG) signal to measure the electrical activity of muscles is classified effectively by ANN [21]. KNN is a simple and powerful classification method helping classify diseases with physiological signals. For examples, Personal system of detecting stress level is developed using physiological signals such as ECG, Chest Expansion, SpO_2 and Electrodermal Activity (EDA) [22]. Using ANN classifier to analyze stockwell transform based on Electroencephalographic (EEG) signal as a non-invasive measurement in classifying mental diseases performs the best among other classifiers [23]. Researches of emotion recognition particularly likely to use physiological signals such as Masseter Muscle (EMG), Blood Volume Pressure (BVP), GSR, RESP, and Respiration Change (RSP) to analysis human emotion [24][25]. Classifying Obstructive Sleep Apnea (OSA) using ECG and SpO_2 obtains good accuracy [26].

From all previous studies, to the best of our knowledge, there is no similar study about classification of CVD patients based on sugar consumption to stress on the controllability. On the other hand, FPG and HbA1c as physiological signals are non-invasive ways to measure blood sugar. Therefore, this research is similar to the mentioned researches where they also used non-invasive physiological signals to classify or detect diseases using ANN and KNN that offered great results. ANN and KNN will be used in this paper as well for classifying elderly group based on sugar consumption using FPG and HbA1c measurements.

III. METHODOLOGY

This paper aims to classify elderly groups who have the potential to control their blood sugar based on sugar consumption. ANN and KNN classifiers are utilized for classifying in accordance with the physiological parameters comprising FPG and HbA1c. The groups of elderly people are separated into controllable group and uncontrollable group. The methodology sequentially discusses the following four steps consisting data collection, data classification, model construction and model verification, respectively.

A. Data Collection

The anonymized data is collected with approval from the university ethics committee. The data used for classification comes from an elderly group from one community in Chiang Rai, Thailand. The dataset includes 53 males and 68 females. The FPG and HbA1c are selected from their health reports. An example of collected data is shown in Table I. Each elderly people checked their health twice, T1 means first-time check, T2 means second-time check after six months.

TABLE I. COLLECTED DATA EXAMPLE

Person. No	Factors			
	T1		T2	
	HbA1c (%)	FPG (mg/dl)	HbA1c (%)	FPG (mg/dl)
1	11.7	189	10.4	175
2	7.4	134	7	123
3	6.9	187	6.9	187
4	11.9	148	11.7	133
5	7.4	133	7.4	111
6	6.3	141	6.5	132
…	…	…	…	…
121	7.4	62	7	78

B. Data classification

Classifying the collected data into groups is the main goal of this step. As mentioned before, the data in T1 shows elderly people's HbA1c and FPG for the first time, doctors provided some recommendations to each of them about sugar consumption after checking their physiological report, then they checked about HbA1c and FPG again at T2. The evaluating criteria are that as long as HbA1c(T1) is greater than HbA1c(T2) and FPG(T2) is smaller than FPG(T1), then the person will be considered into controllable group (C), otherwise uncontrollable group (U). The data classification example is presented in Table II.

TABLE II. CLASSIFIED DATA EXAMPLE

Person. No	Factor				Group
	T1		T2		
	HbA1c (%)	FPG (mg/dl)	HbA1c (%)	FPG (mg/dl)	
1	11.7	189	10.4	175	C
2	7.4	134	7	123	C
3	6.9	187	6.9	187	C
4	11.9	148	11.7	133	C
5	7.4	133	7.4	111	C
6	6.3	141	6.5	132	U
…	…	…	…	…	
121	7.4	62	7	78	U

C. *Model Construction*

ANN and KNN classifier are used for constructing a model after finished classifying collected data into two groups. All data in ANN model will use the cross-validation technique by 70% training data and 30% testing data, while KNN does not need training since KNN just calculates the distance between variables. The description of the two classifiers are shown below:

1) ANN Classifier

ANN as a practical and useful classification method learned from biological learning system and used for solving real-world problems. Its components include interconnected neurons involved in three parts which are an input layer, one or multiple hidden layers and an output layer [27]. As Figure I shows, inputs are known variables, after connecting weight between input layer(n) and one or more hidden layers(m), the bias (b) is followed [27], the two outputs are the classification groups for this paper.

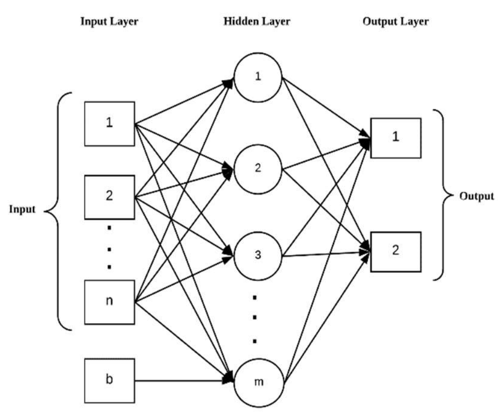

Fig. 1. Artificial Neural Network Conpectual framework.

2) KNN Classifier

KNN was introduced by Fix and Hodges in 1951, it is a nonparametric and supervised learning algorithm, which is one of the most widely used classification methods. K representing the limited number of nearest neighbors plays an important role in the majority vote [28]. The object of KNN is to collect the nearest data with K number restraint. It uses the Euclidean distance formula to calculate between a test sample and specified training samples as presented in equation (1) , x_i is the input sample data, p represents the total features of x_j, and n is the total number of x_i [28].

$$\mathrm{d}(x_i, x_j) = \sqrt{(x_{i1} - x_{j1})^2 + (x_{i2} - x_{j2})^2 + \cdots + (x_{in} - x_{jp})^2} \quad (1)$$

D. *Model Verification*

The final step of the methodology is to verify the accuracies of ANN and KNN classifier. The verification of accuracy is shown in equation (2). By using all correct classifications (TP+TN) to divide by all example data (TP+TN+EP+FN) forms the verified accuracy. TP presents True Positive, TN presents Ture Negative, FP presents False Positive and FN presents False Negative [29]. And the equation of precision and recall for the model are shown in equation (3) and equation (4) respectively.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

$$\text{Precision} = \frac{TP}{(TP + FP)} \quad (3)$$

$$\text{Recall} = \frac{TP}{(TP + FN)} \quad (4)$$

IV. RESULTS AND DISCUSSION

This part demonstrates the classification results using ANN and KNN classifiers on elderly people's data. The accuracy of ANN and KNN with class precision and class recall is presented in Table III and Table IV, they are the matrix of predicted (Pred.) uncontrollable group and true uncontrollable group.

TABLE III. ANN CLASSIFICATION RESULT

	True Uncontrollable Group	True Controllable Group	Class precision
Pred. Uncontrollable Group	120	50	70.59%
Pred. Controllable Group	0	0	0.00%
Class recall	100.00%	0.00%	Accuracy : 70.59%

TABLE IV. KNN CLASSIFICATION RESULT

	True Uncontrollable Group	True Controllable Group	Class precision
Pred. Uncontrollable Group	39	16	70.91%
Pred. Controllable Group	12	5	29.41%
Class recall	76.47%	23.81%	Accuracy :

			61.11%

Both accuracies of ANN and KNN did not achieve good performance. Table III is the result of ANN classification method. The classification accuracy was 70.59% and its class precision and class recall for uncontrollable group showed 70.59% and 100.00% respectively, but its class precision and class recall for controllable group both displayed 0.00%. Table VI shows the classification result of KNN classification method. The KNN accuracy was 61.11% which was lower than ANN classifier. The class precision and class recall of KNN for uncontrollable group were 70.91% and 76.47% respectively, and the class precision and class recall of KNN for controllable group were 29.41% and 23.81% respectively. It can be seen that KNN performed poorly.

Due to the low accuracies of ANN and KNN, the reasons of this phenomenon should be investigated. Firstly, the elderly data is relatively low since the sample size was 242, that could make ANN and KNN predict incorrectly, thus only 4 factors were considered in data classification. Secondly, from the elderly data, the levels of HbA1c and FPG of most senior citizens were already higher than the standard level as mentioned previously, hence their data may be classified wrongly.

Anyhow, from the classification results of ANN and KNN, ANN obtains higher accuracy than KNN, because ANN performed good result on uncontrollable group with relatively high accuracy compared to ANN. As for KNN, K takes the main role in gathering the nearest neighbors, it could affect the accuracy a lot based on determining K, and K has lower accuracy on class precision and class recall. Although KNN performed better on controllable group, the overall accuracy of ANN is higher. Therefore, ANN can be more suitable for classifying elderly group based on sugar consumption.

V. Conclusion

To the best of our knowledge, there is no existing research about classification of elderly group based on sugar consumption to reflect on the controllability of blood sugar. This is a relatively new topic in physiological signals studies. This paper suggests elderly people to control blood sugar based on sugar consumption, and it presents a classification of group involving senior citizens using ANN and KNN classification method. The gathered data of senior citizens was total of 242 from 121 elderly people. It included two variables which are HbA1c and FPG for measuring blood glucose, and it was composed of first-time check and second- time from elderly people. Based on the two variables for two times check, elderly people separated into controllable group and uncontrollable group, which means a person can or cannot control their blood sugar in accordance with classification rules. From the classification result of ANN and KNN classifier, ANN can provide better result for classifying elderly group. The future work will focus on how to improve the accuracy of ANN for elderly people in order to prevent them from excessive sugar consumption.

References

[1] Knodel, J., & Chayovan, N. (2008). Older persons in Thailand: a demographic, social and economic profile. *Ageing international*, *33*(1-4), 3-14.

[2] Nagaratnam, N., Nagaratnam, K., & Cheuk, G. (2016). *Diseases in the Elderly: Age-related Changes and Pathophysiology*. Springer.

[3] World Health Organization (2016) Cardiovascular Disease (CVDs) Fact Sheet. WHO, Geneva.

[4] World Health Organization. (2018). Noncommunicable diseases country profiles 2018.

[5] Suriyawongpaisal, P. (2003). Cardiovascular risk factor levels in urban and rural Thailand-The International Collaborative Study of Cardiovascular Disease in Asia (InterASIA); The InterASIA Collaborative Group∗. *European Journal of Cardiovascular Prevention & Rehabilitation*, *10*(4), 249-257.

[6] Khatab, K., Inthawong, R., Whitfield, M., Collins, K., Raheem, M. A., & Ismail, M. (2019). Application of Cardiovascular disease (CVD) risk assessment equations to the Thai population. *Biostatistics and Epidemiology International Journal*.

[7] Lim, L., Banwell, C., Bain, C., Banks, E., Seubsman, S. A., Kelly, M., ... & Sleigh, A. (2014). Sugar sweetened beverages and weight gain over 4 years in a Thai national cohort–a prospective analysis. *PLoS One*, *9*(5), e95309.

[8] Vapattanawong, P., Aekplakorn, W., Rakchanyaban, U., Prasartkul, P., & Porapakkham, Y. (2010). Obesity and mortality among older Thais: a four year follow up study. *BMC Public Health*, *10*(1), 604.

[9] Johnson, R. K., Appel, L. J., Brands, M., Howard, B. V., Lefevre, M., Lustig, R. H., ... & Wylie-Rosett, J. (2009). Dietary sugars intake and cardiovascular health: a scientific statement from the American Heart Association. *Circulation*, *120*(11), 1011-1020.

[10] Lim, L., Banwell, C., Bain, C., Banks, E., Seubsman, S. A., Kelly, M., ... & Sleigh, A. (2014). Sugar sweetened beverages and weight gain over 4 years in a Thai national cohort–a prospective analysis. *PLoS One*, *9*(5), e95309.

[11] Languepin, O. (2019, February 07). Thailand's dangerous sugar addiction. Retrieved from https://www.thailand-business-news.com/news/headline/51508-thailands-dangerous-sugar-addiction.html

[12] Lustig, R. H., Schmidt, L. A., & Brindis, C. D. (2012). Public health: The toxic truth about sugar. *Nature*, *482*(7383), 27.

[13] Patuwo, E., Hu, M. Y., & Hung, M. S. (1993). Two-group classification using neural networks. *Decision Sciences*, *24*(4), 825-845.

[14] World Health Organization. (2006). Definition and diagnosis of diabetes mellitus and intermediate hyperglycaemia: report of a WHO/IDF consultation.

[15] American Diabetes Association. (2006). Standards of medical care in diabetes-2006. *Diabetes care*, *29*(1), S4.

[16] Kahn, R. (2003). Follow-up report on the diagnsis of diabetes mellitus: the expert committee on the diagnosis and classifications of diabetes mellitus. *Diabetes care, 26*(11), 3160.

[17] Singh, R. R., Conjeti, S., & Banerjee, R. (2013). A comparative evaluation of neural network classifiers for stress level analysis of automotive drivers using physiological signals. *Biomedical Signal Processing and Control*, *8*(6), 740-754.

[18] Faust, O., Martis, R. J., Min, L., Zhong, G. L. Z., & Yu, W. (2013). Automated detection of pulmonary edema and respiratory failure using physiological signals. *Journal of Medical Imaging and Health Informatics*, *3*(3), 424-431.

[19] Acharya, R., Kumar, A., Bhat, P. S., Lim, C. M., Kannathal, N., & Krishnan, S. M. (2004). Classification of cardiac abnormalities using heart rate signals. *Medical and Biological Engineering and Computing*, *42*(3), 288-293.

[20] Nayak, S. G., & Nayak, D. (2012). Classification of ECG signals using ANN with resilient back propagation algorithm. *International Journal of Computer Applications*, *54*(6), 20-23.

[21] Ahsan, M. R., Ibrahimy, M. I., & Khalifa, O. O. (2011, May). Electromygraphy (EMG) signal based hand gesture recognition using artificial neural network (ANN). In *2011 4th International Conference on Mechatronics (ICOM)* (pp. 1-6). IEEE.

[22] Attaran, N. , Brooks, J. , & Mohsenin, T. . (2017). A low-power multi-physiological monitoring processor for stress detection. *Sensors*. IEEE.

[23] Hariharan, M. , Vijean, V. , Sindhu, R. , Divakar, P. , Saidatul, A. , & Yaacob, S. . (2014). Classification of mental tasks using stockwell transform. *Computers & Electrical Engineering, 40*(5), 1741-1749.

[24] Meftah, I. T. , Thanh, N. L. , & Amar, C. B. . (2012). Emotion Recognition Using KNN Classification for User Modeling and Sharing of Affect States. *Proceedings of the 19th international conference on Neural Information Processing - Volume Part I.*

[25] Wagner, J. , Kim, N. J. , & Andre, E. . (2005). From Physiological Signals to Emotions: Implementing and Comparing Selected Methods

for Feature Extraction and Classification. *2005 IEEE International Conference on Multimedia and Expo*. IEEE Computer Society.

[26] Memis, G. , & Sert, M. . (2017). [ieee 2017 ieee 11th international conference on semantic computing (icsc) - san diego, ca, usa (2017.1.30-2017.2.1*)] 2017 ieee 11th international conference on semantic computing (icsc) - multimodal classification of obstructive sleep apnea using feature level fusion.* 85-88.

[27] Yao, X. (1999). Evolving artificial neural networks. *Proceedings of the IEEE*, *87*(9), 1423-1447.

[28] Peterson, L. E. (2009). K-nearest neighbor. Scholarpedia, 4(2), 1883.

[29] Intayoad, W., Becker, T., & Temdee, P. (2017). Social Context-Aware Recommendation for Personalized Online Learning. *Wireless Personal Communications, 97*(1), 163-179.