Stroke Risk Prediction Model

WAN NOR SYUHADA WAN ZAHARI[1], EKO SUPRIYANTO[1,2,3], NAZRIAH MAHMUD[1]
[1] School of Biomedical Engineering and Health Sciences, Faculty of Engineering,
[2]IJN-UTM Cardiovascular Engineering Centre,
[3]Advanced Diagnostics and Progressive Human Care Research Group,
Universiti Teknologi Malaysia, Johor, MALAYSIA
eko@utm.my

Abstract: - The purpose of this study is to improve the existing stroke risk prediction model for the next 10 years. Current existing risk prediction model was done based on the data obtained mostly from America and Africa. There are a few risk prediction models done based on the data from Asian participants. Hence, this paper will predict the risk for stroke for the next 10 years using data collected from Asia. Data from Korean and China risk prediction model were obtained and sorted according to categories. The weightage of each risk factors is then calculated. This study also used the artificial intelligence as a comparison to predict the risk of stroke. Data from a few studies obtained and compared to estimate the probability of stroke that may occur in 10 years of time. Then, a set of data will be used to train the Artificial Neural Network (ANN) and compared to the results obtained using conventional calculation method. The usage of ANN to predict and learn about the risk factors for stroke will give a significant benefit in the future. This study developed a personalized stroke risk prediction which expected to be better and relevant to Asian people compared to other risk prediction model.

Key-Words: - ANN, Prediction, Risk, Stroke

1 Introduction

Stroke is the third leading causes of death in USA as recorded by Stroke Centre USA. Stroke also known as leading causes of death for those age above 65 years old in Korea [1]. Other than that, stroke survivor will incur extra budget in term of health expenditure for treatment such as rehabilitation for post stroke.

In the brain, there is a network of artery which bring blood supply to the tissue to support oxygen and nutrient demand. However, when there is blockage of the blood vessel, the demand cannot be fulfilled and causing the affected tissue to become ischemic. Further delay of supply disruption will cause infarction and then stroke will occur. There are a few factors causing the blockage of the artery which are blood clot (thrombus), fat accumulation at the vessels itself (embolic) and tumour which shift the tissue position and causing pressure against the skull. The tissue will then become infarcted and necrotic thus weakening the body, slurred speech and many other sign and symptoms. Stroke can be diagnosed usually based on the radiology images and biomarkers. Stroke can be fatal due to aspiration, respiratory failure and many others. According to Jee et.al [1], there are three types of stroke namely ischemic stroke, haemorrhagic stroke and transient ischemic attack. However, this study only focuses on ischemic stroke due to its highest stroke case nowadays.

There are few risk factors associated with stroke which are age, gender, body mass index, systolic blood pressure, blood glucose, total cholestrol level, smoking status, alcohol intake, physical activity as well as family history [1-3].

Risk factors are divided into three categories; Low Energy Level, (LEL), Moderate Energy Level, (MEL) and High Energy Level (HEL) as shown in Table 1. LEL are the risk factors that unmodifiable including age, gender, family history, race. While for the MEL, any biomarkers related to the disease, like blood pressure, total cholesterol level, BMI, diabetes. biosignal, electrocradiogram (ECG), and any other symptoms. HEL is the modifiable risk factors, such as smoking habit, fluid intake including alcohol, mental status, physical activity and environment.

Risk factor increased linearly with age probably due to decrease in kidney mass and blood vessels elasticity. These will increase blood pressure and blood clot can be easily formed. Hypertension is the most signifant factor causing stroke. One of the reason is due deacreased of bioavalabality of Nitric Oxide (NO) which then will reduce the endothelial function and will further lead to the endothelial dysfunction. If this happen, artherosclerosis and vascular disesase can be easily happened [4].

Table 1. Classification of Risk

Low Energy Level (Molecular Structure)	**Middle Energy Level** (Body System)	**High Energy Level** (Bioenergy Symphony)
DNA Telomere (Age) Chromosome (Gender) Genetic * Abnormalities * Race * Parent History	Cardiovascular System: * Heart rate (HR) * Blood Pressure - Systolic - Diastolic Body Mass Index Blood Lipid * Total Cholesterol (TC) * LDL * HDL Blood Sugar	Food Fluid Air Environment Physical activity Mental activities

Table 2. Existing Stroke Risk Prediction Model

	Framingham Study	**Diamond Forrester**	**Procam**
Type of risk prediction	Predict risk of cardiovascular events such as coronary heart disease, stroke, peripheral artery disease and heart failure	Predict the risk of having significant coronary artery disease	Predict 10-year risk of coronary events
Risk factors considered	• Sex specific incorporate age • Total and HDL cholesterol • Systolic Blood Pressure (SBP) • Treatment for Hypertension • Smoking status • Diabetes	• Age • Sex • Type of chest pain - Typical - Atypical - Non-angina	• Age • SBP • HDL • Family history of premature MI • Diabetes • Triglyceride level
Classification: Low Intermediate High	 < 10% 10-20% >20%	 <30% 30-70% >70%	 <10% 10-20% >20%

2 Problem Formulation

The previous methods of risk prediction have been developed using existing data and it is static. For example, the Framingham study as discussed in Table 2 was done since 1948. Hence, some data might be obsolete as there might have changes in the epidemiology after that. The objectives of this paper are to develop a new model for stroke prediction and to create a dynamic risk predictor using artificial intelligence. This paper will focus on creating an equation to calculate risk prediction for stroke with existing database and train machine learning to identify the best equation that suit the prediction and will keep changing with the data entries later on. From the study done, it is expected that a better risk prediction which is dynamic, user friendly, fast, accurate predictor can be produced.

.

Risk Factor	Man	Woman
LEL: Age [5]	$y = 2.83 \times 10^{-2}x^2 + 0.14x + 0.0429$	
MEL: Systole [6]	$y = 2.65 \times 10^{-4}x^3 - 1.06 \times 10^{-1}x^2 + 14.18\,x - 611.07$	$y = 9.24 \times 10^{-5}x^3 - 3.17 \times 10^{-2}x^2 + 3.74\,x - 122.26$
MEL: Diastole [5]	$y = -1 \times 10^{-3}x^2 - 0.3385\,x + 967.69$	
MEL: Blood Glucose [7]	$y = 8.72 \times 10^{-3}x^4 - 3.34 \times 10^{-2}x^3 + 4.74\,x^2 - 2.95\,x + 6877.20$	
MEL: Total Cholesterol[7]	$y = 8.26 \times 10^{-3}x^2 - 3.15\,x + 400.74$	$y = 1.07 \times 10^{-2}x^2 - 4.37\,x + 551.10$
MEL: BMI[2]	$y = -7.28 \times 10^{-2}x^4 + 7.36x^3 - 277.97\,x^2 + 4651.14\,x - 29128.85$	$y = -0.21\,x^4 + 20.63\,x^3 - 749.24\,x^2 + 12014.54\,x - 71839.28$

Table 3. Risk factor equation based on graphs

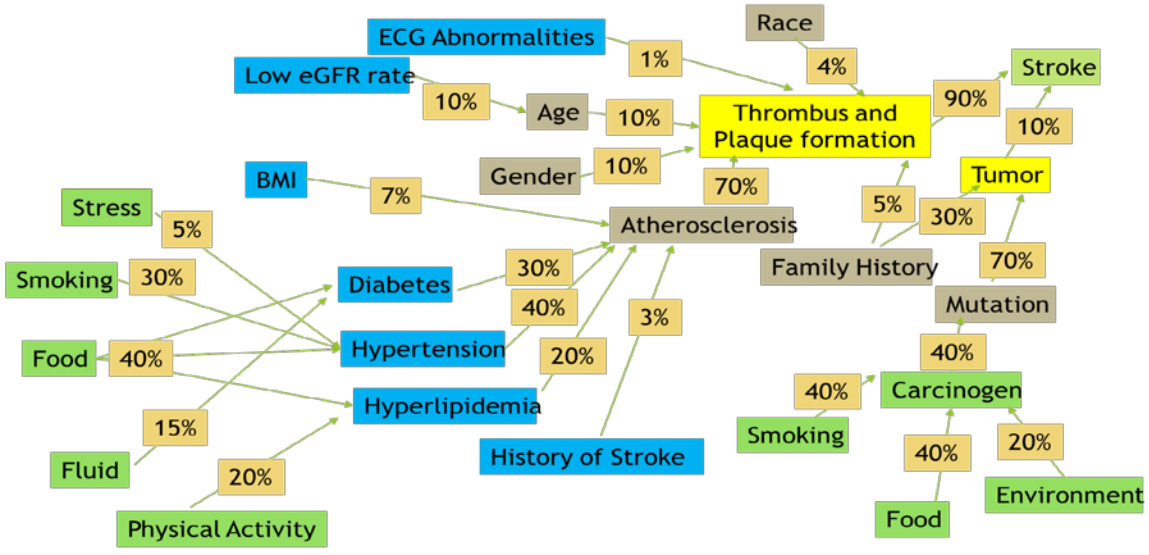

Fig. 1. Decision Tree for Risk Calculation

3 Method and Material

The data for this study is obtained from retrospective study then compared with sample of data from online risk predictor. Retrospective data will be extracted, and form equation as shown in Table 3. Then a decision three as depicts in Fig. 1 is formed in order to give weightage to each of the risk factors

3.1 Basic risk calculation model

As shown in Fig. 1, the decision tree contains the weightage of each risk factors according to the biosymphony energy level. Noted that chocolate colour are the LEL, blue colour are the MEL, and Green are the HEL. From this, probability gained from retrospective data obtained from a few existing study were then compared and calculated using equation (1).

$$\boldsymbol{p} = [a(g * w) + bp(g * w) + dm(g * w) + bmi(g * w) + tc(g * w) + sm(g * w) + ai(g * w) + pa(g * w)]] \quad (1)$$

Where a is age, bp is blood pressure, dm is diabetes, bmi is body mass index, tc is total cholesterol, sm is smoking status, ai is alcohol intake and pa is physical activity.

3.2 Database from literature

Table 4 is the summary of the databases used. These data will be used as the base of the predictor for each risk factors. Table 3 shows the equations based on the graphs.

Table 4. Summary of Databases

Author	Sample Size	Location
S.H. Jee et al.[1]	1, 329, 525	Korea
S.Fukui et al.[9]	400	Japan
J.Y. Al Hashel et al.[10]	984	Kuwait
Chien et al.[11]	49, 281	China
J.W.Lee et al.[12]	6, 885, 789	Korea

3.3 Conventional equation algorithm

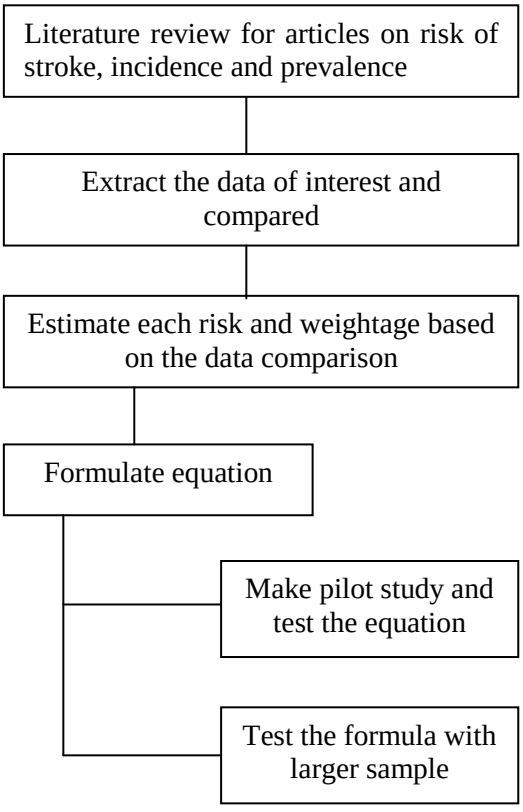

Fig. 2. Flow chart of equation formula

Figure 2 illustrates the flow for conventional equation algorithm process. The process began with literature review from articles on risk of stroke, incidence and prevalence. The data from the articles then extracted and compared. From the data comparison, each risk and weightage are estimated. Equation then formulated and pilot study is conducted to test the equation. A large sample is required to test the formula.

3.4 Machine learning algorithm

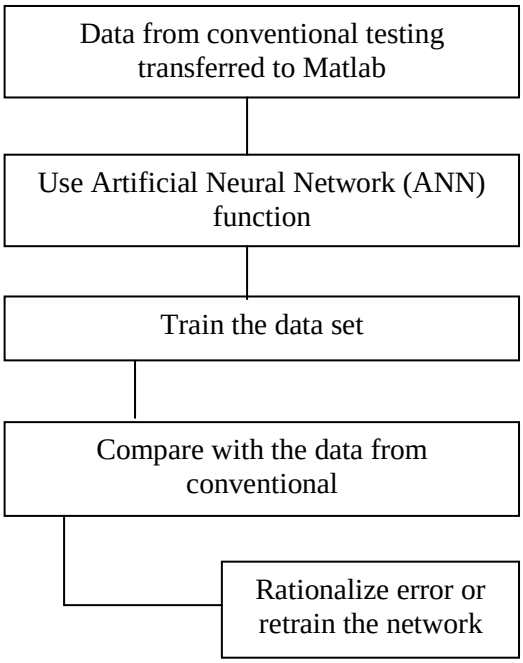

Fig. 3. Flow chart of machine learning algorithm process

Figure 3 shows machine learning algorithm process. Different from conventional equation algorithm, data from conventional testing are transferred to MATLAB. Artificial Neural Network (ANN) is then used to train the data set. The data obtained then compared with the data from conventional equation. Finally, the error is rationalized and retrain the network.

4 Result and Analysis

Fig. 4 shows the percentage of prediction from twenty samples using conventional equation. Only two samples can be classified as high risk while five of them classified as intermediate risk. The rest of thirteen samples are classified as low risk. Fig. 5 shows the percentage of prediction using ANN. From all twenty samples, eight of them are classified as intermediate risk while twelve samples

are classified as low risk. However, there are no data classified as high risk using ANN.

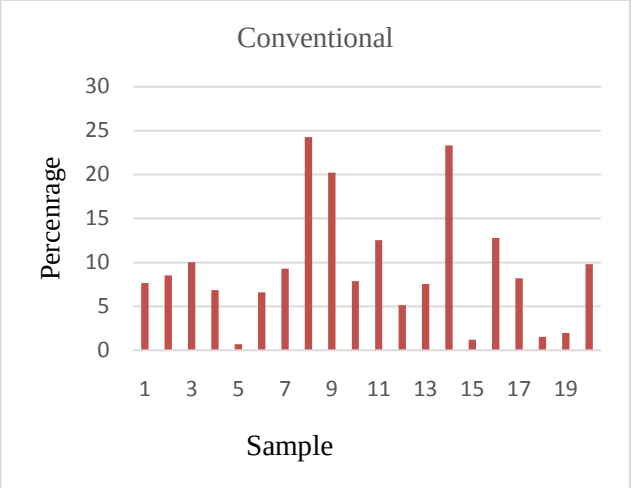

Fig. 4. Result by using conventional equation

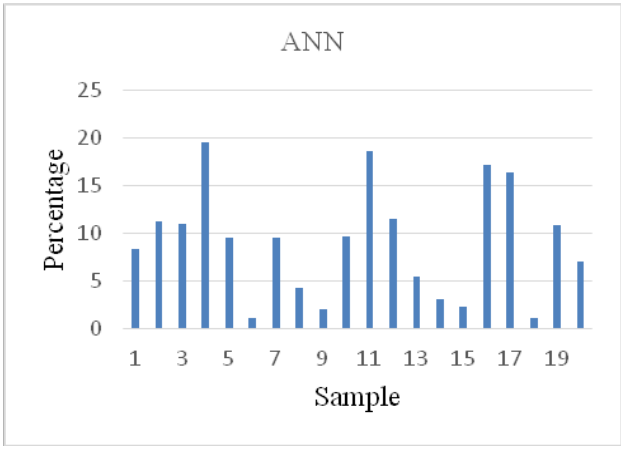

Fig. 5. Result by using machine learning

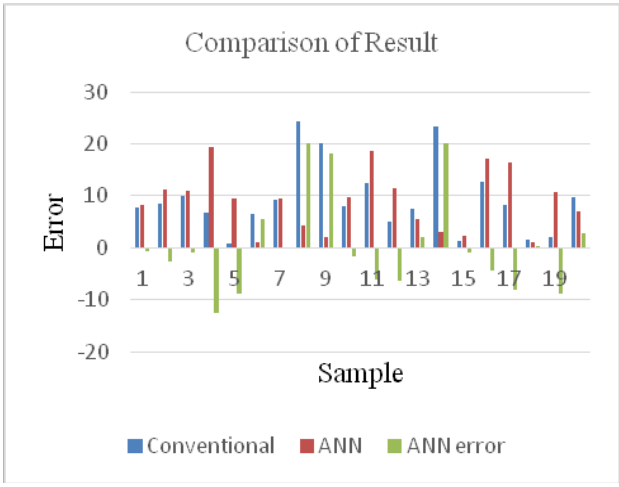

Fig. 6. Comparison result using conventional equation and result from literature

Fig. 6 shows the comparison results between conventional equation and ANN. An error from the comparison is plotted for each sample From the comparison we can see that there are few samples that have a significant different between those two method. First, it is may be due to the number of sample used for training. In the future, larger sample number must be used. Secondly, this may due to the calculated weightage are not accurate. Hence there is the use of artificial intelligence which later it can be changed with increasing database size.

Data obtained usually will separate the data by male and female, hence it is difficult to eliminate underlying factor whether or not the difference is caused by the gender as well. For example, smoking, as we know, women are less smoking, so that risk might be smaller in women compared to men.

As the previous predictor is usually originate from the western, the comparison from Asia may shows some difference. For example, the risk of diabetes is larger in USA compared to Asia. This may due to different lifestyle from both region.

5 Conclusion

After considering multiple databases from previous study, conventional equations are formulated. Since the past few decades, the existing prediction methods are static, but our health risk factors have been changing the trends. For example, previously stroke only attacks the elderly starting from age over 40 years, but now risk of getting stroke has reduced to age of 20 years. So, this method of prediction is the best method where there are dynamic change.

References:

[1] S. H. Jee, J.W. Park, S.Y Lee, et.al, Stroke risk prediction model: A risk profile from the Korean study, *Atherosclerosis*, Vol. 197, No. 1, 2008, pp. 318–325.

[2] J.Lee *et al.*, The development and implementation of stroke risk prediction model in National Health Insurance Service's personal health record, *Computer Methods and Programs in Biomedicine*, Vol. 153, 2018, pp. 253–257.

[3] A. Alloubani, A. Saleh, and I. Abdelhafiz, Hypertension and diabetes mellitus as a predictive risk factors for stroke, *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, Vol. 12, No. 4, 2018, pp.577-584.

[4] Y. Higashi and M. Yoshizumi, Exercise and endothelial function: role of endothelium-

derived nitric oxide and oxidative stress in healthy subjects and hypertensive patients, *Pharmacol. Ther.*, Vol. 102, No. 1, 2004, pp. 87–96.

[5] J. Lam, E. Supriyanto, F. Yahya, et.al, Online Adaptive Coronary Heart Disease Risk Prediction Model, *MATEC Web of Conferences*, Vol. 125, 2017, pp. 1-5.

[6] S.W. Yi, Y. Mok, H. Ohrr, et al., Low Systolic Blood Pressure and Vascular Mortality among More Than 1 Million Korean Adult, *Circulation*, Vol. 133, No. 24, 2016, pp. 2381–2390.

[7] Emerging Risk Factors Collaboration, Diabetes mellitus, fasting blood glucose concentration, and risk of vascular disease: a collaborative meta-analysis of 102 prospective studies, *Lancet*, Vol. 375, no. 9733, 2010, pp. 2215–2222.

[8] S. Fukui, R. Imazeki, Y. Amano, K. Amari, et.al, Common and specific risk factors for ischemic stroke in elderly: Differences based on type of ischemic stroke and aging, *Journal of the Neurological Sciences*, Vol. 380, 2017, pp. 85–91.

[9] J.Y. Al-Hashel, Al-Alya Al-Sabah, S. F. Ahmed, Maha Al-Enezi, Nour Al-Tawheid, Zainab Al Mesailekh, Jasmine Eliwa, Raed Alroughani, Risk Factors, Subtypes, and Outcome of Ischemic Stroke in Kuwait: A National Study, *Journal of Stroke and Cerebrovascular Diseases*, Vol. 26, No.9, 2016, pp. 2145-2152.

[10] K.L. Chien, T.C. Su, H.C. Hsu, W. T. Chang, *et al.*, Constructing the Prediction Model for the Risk of Stroke in a Chinese Population: Report From a Cohort Study in Taiwan, *Stroke*, Vol. 41, No. 9, 2010, pp. 1858–1864.

[11] J.W. Lee, H.S. Lim, D.W. Kim, S.A. Shin, J. Kim, B. Yoo, K.H. Cho, The development and implementation of stroke risk prediction model in Natioanal Health Insurance Service's personal health record, *Journal of Computer Methods and Programs in Biomedicine*, Vol.153, 2018, pp.253-257.