

# Online Adaptive Coronary Heart Disease Risk Prediction Model

*Jostinah Lam<sup>1</sup>, Eko Supriyanto<sup>1,\*</sup>, Faris Yahya<sup>1</sup>, Muhammad Haikal Satria<sup>1</sup>, Suhaini Kadiman<sup>2</sup>, Aizai Azan<sup>2</sup>, and Amiliana Soesanto<sup>3</sup>*

<sup>1</sup> Advanced Diagnostics and Progressive Human Care Research Group, IJN-UTM Cardiovascular Engineering Centre, Faculty of Biosciences and Medical Engineering, Universiti Teknologi Malaysia, 81310 Skudai, Malaysia.

<sup>2</sup> National Heart Institute, Kuala Lumpur, Malaysia

<sup>3</sup> National Cardiovascular Centre, Jakarta, Indonesia

**Abstract.** Coronary Heart Disease (CHD) is the leading causes of death worldwide. Life style changing is one of the important methods to delay the incidence of CHD. The awareness of life style changing is however still low. In order to improve awareness of life style changing, some CHD risk prediction models have been introduced. The existing models however either not well structured, not completed, static or offline. This paper introduces a new online CHD risk prediction model. The model is structured according to three risk factor groups including molecular structure, body system vital sign and bioenergy symphony. The model had also been compared with 5 existing models. Comparison results show that the model has better structure, adaptability and accessibility. Validation test using 120 subjects shows that the model prediction accuracy is 96.2%. This shows that the model is suitable to be used widely for CHD risk prediction both healthy and risk subjects as a preventive method in getting CHD in the earlier age.

## 1 Introduction

Coronary heart disease according to WHO are diseases that related to blood vessels supplying the heart muscle [1]. Coronary heart disease (CHD) is usually caused by blockage that stops the process of supplying blood to the heart. The main reason for the blockage formation is due to the buildup of plaque in the walls of coronary arteries over the years. CHD events can be specified including stable and unstable angina, acute myocardial infarction and coronary death [2].

In 2012, out of the 17.5 million (31% of all global deaths) cardiovascular deaths, 7.4 million people died from CHD [1]. In Malaysia, CHD is the leading cause of death where it contributes 23.10% of total deaths [3]. The high percentage has put Malaysia in the 33 rank in the world [3]. The worrying numbers of deaths from CHD in Malaysia has alarmed the country to put more efforts on the preventive measures. A number of preventive measures had been carried out through modifiable risk factors such as smoking, hypertension, cholesterol level, diabetes, obesity, diet and physical activity [4]. One of the preventive measures taken is to predict the chances of an individual of getting CHD with the help of risk predictor. By knowing the probability of one having CHD, further actions can be carried out to prevent the occurrence of the disease. Hence, early detection and prevention of CHD is vital in reducing the number of CHD deaths.

There are a variety of risk prediction models throughout the advancement of cardiovascular risk assessment research. One of the most well-known and long-standing risk prediction models would be the Framingham risk score [5]. Nearly all of the risk prediction models were developed based on multiple variables since the happening of CHD event does not depend on solely one but many risk factors. The prediction variables used in the original Framingham CHD risk score include age, gender, total cholesterol, HDL cholesterol, blood pressure, smoking status and diabetes mellitus. The other risk prediction models like SCORE [6], QRISK [7], Reynolds [8], ACC/AHA risk calculator [9] have more or less similar prediction variables with each other [10]. Nevertheless, the existing risk predictors are not well-structured. All of the risk factors are not properly classified into specific category which makes it difficult to track and understand the root of the CHD event. Without a thorough understanding of CHD event and systematic approach to educate public about CHD event, it is challenging to increase public's awareness and hence the difficulty to reduce CHD mortality rate.

Another weakness found in the existing risk prediction model is the elasticity of the model. Most of the risk prediction models use data collected in the past which can be as short as 1 to 3 years ago or as long as 10 to 20 years ago. By the time the prediction model is ready to use, the predicted risk may not be suitable for individuals at that time anymore since the weight of risk

\* Corresponding author: eko@utm.my

factor changes over time. Consequently, there is a need to develop a more dynamic risk prediction model to ensure the flexibility and adaptive of the model in future. Therefore, this study takes the initiative to develop a novel online adaptive CHD risk prediction model. In order to improve trust and increase awareness of people to live on healthier lifestyle, a new risk factor structure based on energy level classification was incorporated in this prediction model. The prediction model was also constructed based on the concept of being adaptive. This adaptive prediction model includes dynamic database, adaptive weighted risk factor and dynamic interpolation.

## 2 CHD risk factor

### 2.1. Physiology of CHD

There are many factors that influence the occurrence of CHD event. All of the risk factors are interrelated with each other and cannot be ruled out. A physiological diagram of CHD was illustrated in Fig. 1. The fundamental of CHD event is that coronary arteries was blocked by plague buildup for years and disable the oxygen supply process to the heart muscle. Therefore, the diagram starts with cardiovascular system in the middle and DNA cell in the centre symbolizes the molecule structure that forms the blood, vessel and heart. The molecular structure carries the DNA of an individual and it is not able to be changed for the whole life. Age, gender and genetic are risk factors that are non-modifiable.

The primary component that causes the buildup plague is cholesterol; hence measurement of total cholesterol, HDL and LDL is very important in determining CHD risk. Other cardiovascular related risk factors like blood pressure, heart rate, and blood sugar is definitely significant in predicting CHD incidence. Besides cardiovascular system, the whole body system is actually involving in the CHD events indirectly. It is the bioenergy symphony that manipulates the operation of body system. As we can see from Fig. 1, food, fluid, air, mental activity, environment and physical activity influence the operation of specific body system. These bioenergy symphonies can influence the body system either positively or negatively. Therefore, people can gain benefits from knowing the physiological of CHD incidence and make a change towards healthy lifestyle.

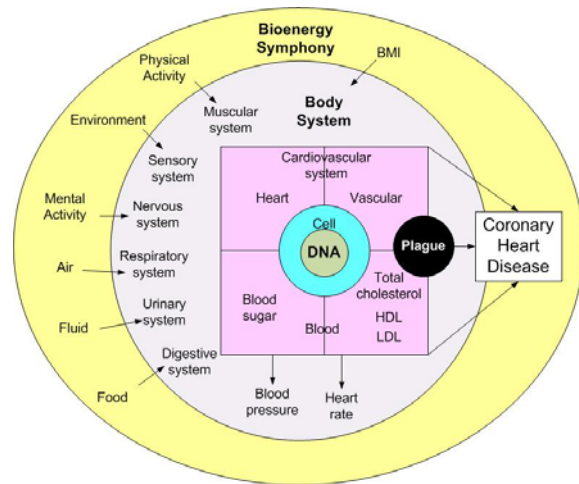


Fig. 1. Physiological diagram of CHD.

### 2.2 Energy Level Classification of Risk Factor

Based on the physiology of CHD event, the risk factors involved can be classified into 3 categories which are molecular structure, body system, and bioenergy symphony. These 3 classes can also be sorted out according to their energy level: low, middle, and high energy level. The energy level indicates the amount of energy required to function in the universe.

The first class of risk factors is molecular structure class which consists of non-modifiable risk factors like age, gender, and heredity. Non-modifiable risk factors can be linked directly to DNA and genetic of an individual where the risk factors in this category cannot be altered anymore. This fundamental molecular structure uses very low energy level to operate.

Next, the second class of risk factor would be body system class where it consists of every related body systems like cardiovascular system, respiratory system, digestive system to CHD. The risk factors in this category indirectly associated to the body system of an individual.

The third class is the high energy level category: bioenergy symphony. Bioenergy symphony class associate with the entire universe. Risk factors that fall on this category are food, fluid, air, environment, physical activity, and mental activity. This category is a significant class in determining the risk factor as all of these risk factors can either increase or decrease the possibility of one in getting CHD depending on their lifestyles. Table 1 shows classification of CHD risk factors.

**Table 1.** Classification of Risk Factor.

Risk Factors		
Low Energy Level (LEL): Molecular Structure	Middle Energy Level (MEL): Body System	High Energy Level (HEL): Bioenergy Symphony
<ul style="list-style-type: none"> <li>DNA Telomere (Age/A)</li> <li>Chromosome (Gender/G)</li> <li>Genetic (N) (Abnormalities, Race, Parent history)</li> </ul>	<ul style="list-style-type: none"> <li>Cardiovascular system (Heart rate/HR, systolic (SB) &amp; diastolic (DB) blood pressure)</li> <li>Body Mass Index (BM)</li> <li>Blood lipid (Total cholesterol/TC, LDL, HDL)</li> <li>Blood sugar (BS)</li> </ul>	<ul style="list-style-type: none"> <li>Food (FD)</li> <li>Fluid (FL)</li> <li>Air (AR)</li> <li>Environment (EN)</li> <li>Physical activity (PA)</li> <li>Mental activity (MA)</li> </ul>

### 3 CHD prediction model

CHD risk is currently calculated using static risk equation as follow:

$$CHDRisk_{static} = f(LEL) + f(MEL) + f(HEL) \quad (1)$$

In order to improve the accuracy of the prediction, a new dynamic risk equation (CHDRisk<sub>dynamic</sub>) is proposed with considering region (R) and time (T).

$$CHDRisk_{dynamic} =$$

$$f(LEL, R, T) + f(MEL, R, T) + f(HEL, R, T) \quad (2)$$

Where:

$$f(LEL, R, T) = (a_0 + a_1A_{G,R,T} + a_2A_{G,R,T}^2)a_3N \quad (3)$$

$$f(MEL, R, T) = f(SB) + f(DB) + f(TC) +$$

$$f(HDL) + f(LDL) + f(BM) + f(BS) \quad (4)$$

Where:

$$f(DB) = a_4DB^3 + a_5DB^2 + a_6DB + a_7 ;$$

$$f(SB) = a_8SB^3 + a_9SB^2 + a_{10}SB + a_{11} ;$$

$$f(TC) = a_{12}TC + a_{13} ; f(HDL) = a_{14}HDL + a_{15} ;$$

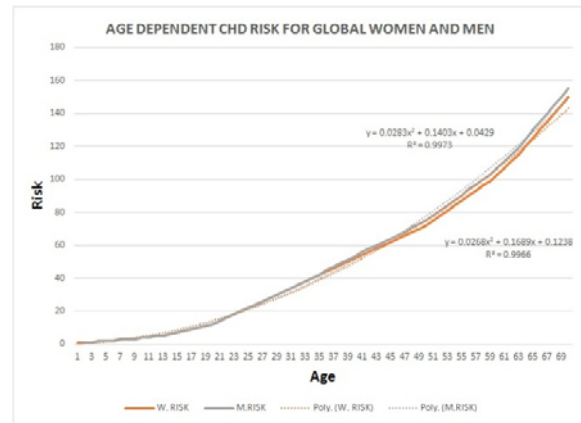
$$f(LDL) = a_{16}LDL + a_{17} ;$$

$$f(BM) = a_{18}BM^3 + a_{19}BM^2 + a_{20}BM + a_{21} ;$$

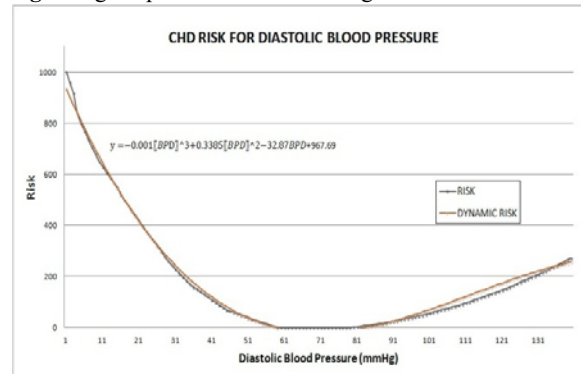
$$f(BS) = a_{22}BM^4 + a_{23}BM^3 + a_{24}BM^2 + a_{25}BM + a_{26} ;$$

$$f(HEL, R, T) = f(FD) + f(FL) + f(AR) + f(EN) + f(PA) + f(MA) \quad (5)$$

The coefficient  $a_n$  can be obtained from dynamic interpolation of epidemiological and individual data. Fig. 2, Fig. 3, and Fig. 4 show the examples of CHD risk for each energy level and its interpolation equation respectively.

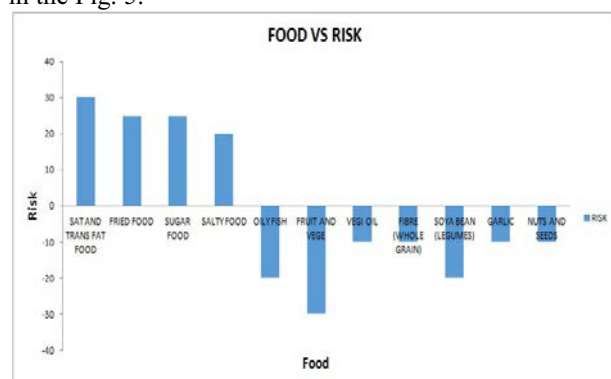


**Fig. 2.** Age dependent CHD risk for global men and women.



**Fig. 3.** CHD risk for diastolic blood pressure.

These graphs were generated from 10900 people with 3 different race groups (Caucasian, Mongolian and Negroid) [2]. This epidemiological data and interpolation equation were inputted in the webserver. This model is used as first condition or static model. In order to calculate the risk, a subject can input the risk factors data via mobile user interface. After risk calculation by the equation in the webserver, the subject can get the information about the current risk, the risk at certain age, the age at 100% risk and list of bioenergy symphony activities to increase age at 100% risk. In addition, the admin can always update the database using new epidemiological data to ensure that the risk is up to date. The model implementation algorithm can be seen in the Fig. 5.



**Fig. 4.** CHD risk for food.

The backend system was done by using MySQL as main database. The web server for front end was done

with Apache 2.0 and coded with PHP version 5.5.1. In order to improve the accuracy of the model to be used in Malaysia, 2410 CHD patients' data from Malaysia was inputted in the model. This input has changed the equation coefficients in the model and the model become dynamic model.

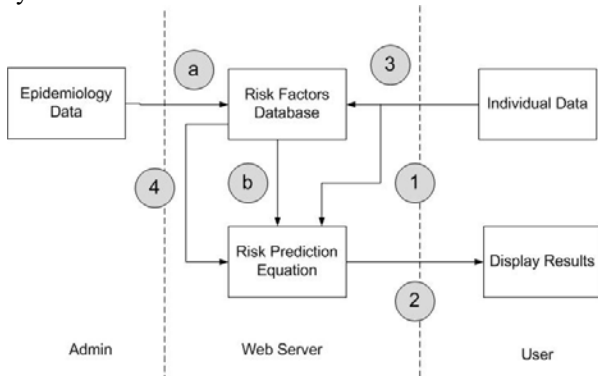


Fig. 5. Model Implementation Algorithm.

### 4 Result and analysis

In order to verify the newly developed adaptive risk prediction model, we did comparison study with 5 existing models including Framingham Risk Score [5], QRISK [7], SCORE [6], Reynolds [8], and WHO/ISH [11]. The comparison includes structure, adaptability and accessibility. The structure analysis is based on criteria of physiological pathway including: completeness, redundancy, connectivity, time analysis and spatial analysis. As for the adaptability analysis, it is conducted based on adaptive weighted risk factor and epidemiology dependent risk calculation. The accessibility analysis includes data enter human error, personalized risk, accountability and real time access.

Table 2 shows the results of structure test carried out with 5 different risk prediction models and one static model. As we can see from Table 2, our dynamic model obtained the highest score among the other existing models. The main criterion that influences the results is the adaptability of the model. The other models scores lower than dynamic model since they have inadequate adaptive weighted risk factor and epidemiology dependent risk calculation. The completeness of the dynamic model also increased the total score in the structure test. However, there are redundancies of structure in the dynamic model because there are some factors that might exist more than one time to be used for different risk factors category. Overall, the adaptive model is up to 41% better than 5 existing models in terms of structure, adaptability and accessibility.

Table 2. Physiological diagram of CHD.

	FRS	QRISK	SCORE	Rey	WHO	Static	Dynamic
Structure	3.6	4.1	3.5	3.9	3.6	3.6	4.7
Adaptability	0.6	0.6	0.6	0.6	0.6	0.6	2
Accessibility	2.3	2.3	2.3	2	2	2	3
Total Point	6.5	7	6.1	6.8	5.7	6.6	9.7
Percentage Different (%)	33	28	37	30	41	32	0

The risk prediction results have also been validated with Electrocardiography and Echocardiography stress test results. There are 120 subjects with age from 15 to 95 years old participated in the validation study. The results from five existing models, static model and dynamic model are shown in Table 3.

Table 3. a) Results of 120 Subjects with existing models.

	FRS	QRISK	SCORE	Rey	WHO	Static	Dynamic
VLR <sup>a</sup>	22	34	18	14	15	23	32
LR <sup>b</sup>	29	26	41	32	46	27	22
MR <sup>c</sup>	43	51	55	42	25	44	46
HR <sup>d</sup>	20	9	4	32	18	19	16
VHR <sup>e</sup>	6	0	2	0	16	7	4

a. VLR = very low risk b. LR = low risk c. MR = middle risk  
d. HR = high risk e. VHR = very high risk

Table 3. b) Results of 120 Subjects with ECG and ECHO tests

	Dynamic	ECG	ECHO	AVG(ECG+ECG)
VLR	32	28	31	29.5
LR	22	23	20	21.5
MR	46	50	47	48.5
HR	16	15	18	16.5
VHR	4	4	4	4

Table 3. c) Accuracy Results

	FRS	QRISK	SCORE	Rey	WHO	Static	Dynamic
VLR	74.57	84.75	61.02	50.85	47.48	77.96	92.19
LR	74.14	79.07	9.30	13.95	51.16	79.63	97.73
MR	88.66	94.85	86.60	51.55	86.60	90.72	94.57
HR	78.79	54.55	24.24	90.91	6.06	86.84	96.88
VHR	66.67	0	50	25	0	57.14	100
AVG	76.57	62.64	46.23	46.45	38.26	78.46	96.27

The risk level was divided into five categories: very low risk, low risk, middle risk, high risk, and very high risk. All of the subjects were also tested with ECG stress test and Echo stress test. Table 3. a) and b) shows the risk prediction results of 120 subjects with existing models and ECG and Echo stress tests. The average of ECG and Echo stress test were then calculated and compared with the other risk prediction models. The accuracy results were shown in Table 3. c). The result of accuracy test show that the prediction accuracy for dynamic model is 96.2%.

Although the model had shown promising results from verification and validation test, this model still has some weaknesses that need to be confronted. First, this model requires bigger resources for its database and processing to produce dynamic equation. Another disadvantage is that there is a possibility for dynamic interpolation error to happen because the new data group has different trend or far distance from existing equation. For example, the new data group comes from huge nuclear radiated area; this will change the trend drastically and introduce error in the interpolation. Generalization of data will also produce inaccuracy if the previous epidemiology data is originated from a specific region. Therefore, in order to improve the model, artificial intelligence can be applied for the interpolation or risk calculation.

## 5 Conclusion

In this study, a new online adaptive CHD risk prediction model had been developed. The model had been developed based on 10900 patient data from available global data and 2410 patients' data from Malaysia. This model had been tested by 120 subjects and validated with ECG and ECHO stress tests. Test result shows that the model just show error less than 4% compared with result from ECG and ECHO stress tests. When compared to other existing models, the model had shown up to 41% better results in term of structure, adaptability and accessibility.

## Acknowledgement

The authors would like to thank Ministry of Education Malaysia for the providing financial support of this project. The authors would also like to express gratitude to National Heart Institute, Malaysia and National Cardiovascular Centre, Indonesia, for the data support.

## References

1. <http://www.who.int/mediacentre/factsheets/fs317/en>
2. N. D. Wong, Nat. Rev. Cardiol., **11.5**, pp. 276-289 (2014)
3. <http://www.worldlifeexpectancy.com/country-health-profile/malaysia/>
4. S. M. Grundy, R. Pasternak, P. Greenland, S. Smith and V. Fuster., J. Am. Coll. Cardiol., **34.4**, pp. 1348-1359 (1999)
5. P. W. Wilson, et al., Circulation, **97.18**, pp. 1837-1847. (1998)
6. I. Graham, et al., Eur. Heart J., **28.19**, pp.2375-2414 (2007)
7. J. Hippisley-Cox, et al., BMJ **335.7611**, pp. 136 (2007)
8. P.M. Ridker, et al., Circulation, **118.22**, pp. 2243-2251 (2008)
9. D.C. Goff, et al., J. Am. Coll. Cardiol. (2014)
10. P.W.F. Wilson, and B. F. Culleton., UpToDate Inc. *Textbook of Medicine* (2010)
11. S. Mendis, et al., J. Hypertens., **25.8** , pp. 1578-1582 (2007)