ISSN: 2141 – 3290
www.wojast.com

UNSUPERVISED CARDIOVASCULAR RISK FACTORS CLUSTERING: TOWARDS AN EXPERT RECOMMENDATION SYSTEM FOR PERSONALIZED NUTRITION THERAPY

[1]EDOHO, M. E.,[1]EKPENYONG, M. E. AND [2]EKONG, A. I.

[1]Department of Computer Science, University of Uyo, Nigeria
mercyedoho@uniuyo.edu.ng, mosesekpenyong@uniuyo.edu.ng
[2]Department of Medical Science and Public Health, Bournemouth University, United Kingdom
aekong@bournemouth.ac.uk

ABSTRACT

High blood pressure with its high prevalence is associated with strong evidence for causing cardiovascular disease (CVD) among other risk factors. Between 1975 and 2015, The World Health Organisation recorded an increase of about 66% in the occurrence of high blood pressure, hence further threatening a reduction in the average life expectancy if left unchecked. This paper analyses cardiovascular risk prediction models and proposes an expert recommendation system for personalised nutrition therapy. The methodology of this study adopts a simulation approach and aims at establishing the feasibility of an ongoing project. Hence, data to the study was generated by imposing a reverse engineering approach on the Framingham's criteria/model using Monte Carlo simulation. An unsupervised learning and correlation hunting of 10,000 patients' cohorts revealed the defect of the Framingham's model, as it failed to establish relationships between other risk factors, save the age factor. Hence, making it an inappropriate model for tackling CVD. To provide efficient prediction of CVD and advance the growing field of healthcare informatics, an expert system framework with 4 components is proposed. The implementation workflow begins with a collection of primary CVD risk factors using the WHO stepwise approach for non-communicable disease surveillance, to offer an effectual means of developing data collection instrument for patients' cohorts data collection. The collected data is then processed into the CVD risk factors database. A fuzzy inference system is deployed to infer appropriate CVD risk scores by formulating a mapping from the input space to output crisp value. These scores are used to label the CVD dataset, for supervised learning and prediction. The predicted CVD risk level is finally matched with suitable diet plan in the nutrition database crafted by dieticians, to provide decision support on personalised nutrition, for persons living with CVD.

INTRODUCTION

Hypertension, a cardiovascular disease, is a morbidity that occurs when a person's blood pressure is raised beyond normal. The measurement of blood pressure considers the rate at which blood flows through the blood vessels and the rate of resistance experienced in the circulatory system. This rate of resistance known as systemic/peripheral vascular resistance creates blood pressure. A rise in the peripheral vascular resistance increases the blood pressure and vice versa (Delong and Sharma, 2020).Measured in millimetres of mercury (mmHg), high blood pressure is defined as systolic blood pressure of at least 140mmHg and diastolic pressure of at most 90mmHg for adults of 18 years and above (Nguyen *et al*., 2013). Systolic blood pressure is the top numeric of the blood pressure measurement; it is the amount of pressure exerted in the blood vessel (arteries) as the heart muscle contracts during the process of pumping oxygen-rich blood into the blood vessel. Diastolic blood pressure is the lower numeric pressure on the blood vessel when the heart muscle is relaxed. Different categories of blood pressure have different systolic and diastolic measurements. When blood pressure falls between 120 and 129/<80mmHg; it is known as elevated blood pressure. Those between 130/80mmHg and 139/89mmHg are categorised as prehypertension. A normal blood pressure (normotension)

lies between systolic pressure less than 120mmHg and diastolic pressure less than 80mmHg. However, in 2017, The American College of Cardiology/American Heart Association revised the categories and guidelines for the detection, evaluation, prevention, and management of adult hypertension. This guideline categorises blood pressure into normal, elevated, stage 1 or 2 hypertension and further defines the threshold of 130/80mmHg for the diagnosis of hypertension (Muntner *et al.*, 2019).

Diverse nutrition research works have been conducted and results have exhibited the corollary of hypertension due to diet. Although the precise cause of primary hypertension remains unidentified, there are several factors that exacerbate its pathogenesis. These factors include over or under expressed gene (Luft, 1998), obesity, high alcohol consumption, high consumption of salt, age, sedentary lifestyle, insufficient potassium, and calcium intake (Sever and Poulter, 1989). Possible management and treatment of idiopathic hypertension involves medication, lifestyle, and nutritional therapy. Nutritional factors play a vital role in either increasing or decreasing the pressure of blood.

Despite the burden of hypertension, most people suffering from this ailment are unaware as the ailment does not present any peculiar symptom(s). However, if left uncontrolled or poorly managed, complications could lead to cardiovascular diseases such as heart failure, stroke (when abnormal and disproportionate pressure hardens the arteries and decreases the rate of blood and oxygen flow to the heart and brain), renal failure, sight impairment, dementia, retinal haemorrhage, and peripheral vascular diseases. A study by Egan (2018) conducted on Coronary Artery Risk Development in Young Adults (CARDIA), observed that the blood pressure of the subjects when they were between the ages of 18 and 30 measured <130/<80 mmHg, but when they turned 55 years, approximately 75% of black men and women had raised blood pressure (BP), compared to an observed variation of 54% of white men and 40% of white women with raised BP. The study also inferred a high rate of hypertension vulnerability among young and middle-aged black adults. According to WHO, 51% of deaths caused by stroke and 45% of deaths due to coronary heart disease stem from high-risk complications of hypertension. Moreover, hypertension increased from 594 million in 1975 to 1.13 billion in 2015 with high prevalence in Africa (27%) and low prevalence in America (18%). While prevalence is low in high income countries, the prevalence in low- and medium-income countries is on the increase. This increase owes to the unawareness of the disease and the underlying risk factors amongst the population (Raji, 2017). However, the state of healthcare in a country constitutes a risk factor as well. An individual with a raised BP in a developed country with excellent treatment and rehabilitation services stands a better chance of disease management and recuperation than her counterpart in low- and middle-income countries with deplorable state of healthcare.

In Nigeria, hypertension is a silent epidemic that is gradually consuming the society. A systematic study by Adeloye *et al.* (2015), conducted using data from Medline, EMBASE and Global health between January 1980 and September 2013 rates 28% hypertension prevalence in Nigeria with distribution of 29.5% and 25% among men and women, respectively, indicating high prevalence in men. In addition, the study showed community estimate of 30.6% prevalence among urban inhabitants and 26.4% among rural inhabitants and further projected about 30.1 million cases of hypertension by 2030.

Recently, therapeutic dietary recommendation systems based on dietary approach to stop hypertension (DASH) have been personalised for weight loss and different chronic diseases. However, drastic modification to one's diet presents a significant change that may be difficult to maintain. Several attempts by most hypertensive patients to adhere to this dietary recommendation seem unyielding. In a cuisine rich country like Nigeria, the abrupt shift from delectable traditional delicacies to consumption of mostly fruits and vegetables tend to be an unrealistic approach that may not be maintained as a lifestyle. Consequently, this work propounds a proactive and highly scalable approach to support hypertension self-management and maintenance of individual's well-being. Medically, physicians routinely make diagnosis of

a condition based on physical examination or answers to the questions provided by the patients before offering either therapeutic prescription or referral to a nutritionist for medical nutrition therapy if the underlying ailment requires nutrition intervention. This approach can be emulated using machine learning and soft computing techniques. Whilst there are various studies on the classification of hypertension risk, little attention has been given to the utilisation of machine learning in the prediction of hypertension risk and its associated medical nutrition therapy. Hence, this paper proposes a nutrition recommendation framework that integrates machine learning, for personalised nutrition therapy. The specific objectives of the paper include, to:

- study frequently used cardiovascular risk models, for easy cohorts' characterisation and system integration,
- simulate a dataset of patients' cohorts using the Framingham's risk criteria,
- learn evolving risk patterns from simulated cohorts using unsupervised machine learning technique.
- propose a recommendation system for matching the predicted risk level with recommended diet plan, for driving decision support systems, towards personalised nutrition therapy.

CARDIOVASCULAR RISK PREDICTION MODELS

Cardiovascular risk predication model is a mathematical equation that evaluates the likelihood of developing CVD within a specified time frame, considering the associated risk factors of the individual (Moons *et al.*, 2012a). The use of risk stratification model in estimating the latent risk of developing atherosclerotic cardiovascular disease (ACVD) has been endorsed globally, as the foremost step in administering the primary preventive therapy (Anderson et al., 2016; Pylypchuk *et al*., 2018). This serves as a methodological strategy that complements the level of risk to the potency of preventive and control therapies. Amongst the myriad of CVD risk prediction models, the most widely recommended is the Framingham risk score (FRS) and pooled cohort equation (PCE). While the guideline of Canadian Cardiovascular Society (CCS) commends the use of FRS in accessing the risk of CVD (Anderson *et al*, 2016), the American College of Cardiology/American Heart Association (ACC/AHA) advocates the use of PCE (Stone *et al*., 2014; Grundy *et al*, 2018). However, there has been conflicting corroborations about the suitability of a CVD risk score in capturing the disparity due to ethics and socioeconomic factors (Pylypchuk *et al*., 2018). The efficiency of a CVD predictive model is reliant on the specific data sample used in the development of the model (Yutsuya, 2018) and the estimation of CVD risk is dependent on the adopted risk model (Boateng *et al*., 2018). According to Ko *et al*. (2020), the FRS and PCE produced high estimates against the actual risk of ACVD in a large population from Ontario, even though PCE was developed and authenticated with sample data from Caucasians and African American men and women with no overt ACVD (Goff et al., 2014). The performance of a CVD risk algorithm therefore differs with different set of data obtained from different populations (Boateng *et al*., 2018).

Despite the application of the CVD risk estimation models in predicting CVD risk in different populations of Sub-Saharan Africa and some regions in Nigeria, no population-based study has been established in Nigeria for the development of CVD risk algorithm for its citizens. Accordingly, one of the objectives of this study is to develop a machine learning model that can predict CVD risk in Nigeria's population and calibrate the model against popular existing CVD risk prediction models.

Framingham Risk Score

According to Gallestery (2016), this gender-specific algorithm was designed and developed for the prediction of 10-year CVD risks of individuals with no obvious indication of heart disease. The algorithm is derived based on the data from the Framingham's heart study. The study which was initiated in 1948 comprised initially of two-thirds (more than 5,200) of the adult residents in Framingham, Massachusetts between the ages of 30 and 62 years. Detailed questions about a participant's lifestyle and blood pressure reading were obtained at interval of 2 years for a period of 14 years. As the study progressed, they kept track of the participants

who developed CVD and those who did not. Over the years of the study, new participants referred to as offspring cohorts were added to the study. In 1971, 5,120 adult children of the participants in the initial cohort and their spouses got added. The third-generation cohort comprised of participants with at least one of the parents in the previous study cohort added in 2001. The maiden version of the risk score algorithm was published in 1998, updated version in 2002 and the current version in 2008. The earliest version included age, gender, LDL cholesterol, HDL cholesterol, blood pressure, diabetes, and smoking. Subsequent versions were updated with dyslipidaemia, age range, hypertension treatment, total cholesterol but excluded diabetes. The presence of diabetes was synonymous to an apparent CHD risk. Table 1 documents the Framingham risk scores for male patients

Table 1. Framingham risk score for male patients

Variable	Range		Point
	Age Range	Variable Measurement	
Age	20-34	-	-9
	35-39	-	-4
	40-44	-	0
	45-49	-	3
	50-54	-	6
	55-59	-	8
	60-64	-	10
	65-69	-	11
	70-74	-	12
	75-79	-	13
Total cholesterol (mg/dL)	20-39	<160	0
		160-199	4
		200-239	7
		240-279	9
		>=280	11
	40-49	<160	0
		160-199	3
		200-239	5
		240-279	6
		>=280	8
	50-59	<160	0
		160-199	2
		200-239	3
		240-279	4
		>=280	5
	60-69	<160	0
		160-199	1
		200-239	1
		240-279	2
		>=280	3
	70-79	<160	0
		160-199	0
		200-239	0
		240-279	1
		>=280	1
Cigarette smoker	20-39	-	8
	40-49	-	5
	50-59	-	3
	60-69	-	1
	70-79	-	1
All non-smokers	-	-	0
HDL cholesterol (mg/dL)	-	<40	2
		40-49	1
		50-59	0
		>=60	-1
Systolic bp (mmHg) - Untreated	Untreated	<120	0
		120-129	0
		130-139	1
		140-159	1
		>=160	2
Systolic bp (mmHg) - treated	Treated	<120	0
		120-129	1
		130-139	2
		140-159	2
		>=160	3

Pooled Cohort Equation

The Pooled Cohort Equation(PCE) which was initiated in 2013 but launched in 2014 by the American College of Cardiology/American Heart Foundation, focuses on the risk estimation of stroke and CHD among those that presents no notable symptom of CVD. This algorithm was introduced due to certain shortcomings experienced with the FRS; the abundance of Caucasians but limited number of other races making up the study cohort and inadequate consideration of ischemic stroke (Preiss and Kristensen, 2015), since stroke may occur before the manifestation of CHD (Stone and Floyd-Jones, 2016). The development of the PCE is based on the collection and analysis of population-based studies (the initial and offspring study cohorts of Framingham heart study, coronary artery risk development in young adults' study) administered with National Heart, Lung and Blood Institute funding. The risk factors examined included age, sex, HDL cholesterol, systolic blood pressure, cholesterol, antihypertensive drug taken, race, smoking status and diabetes status.

Table 2. Framingham risk score for female patients

Variable	Range		Point
	Age Range	Variable Measurement	
Age	20-34	-	-7
	35-39	-	-3
	40-44	-	0
	45-49	-	3
	50-54	-	6
	55-59	-	8
	60-64	-	10
	65-69	-	12
	70-74	-	14
	75-79	-	16
Total cholesterol (mg/dL)	20-39	<160	0
		160-199	4
		200-239	8
		240-279	11
		>=280	13
	40-49	<160	0
		160-199	3
		200-239	6
		240-279	8
		>=280	10
	50-59	<160	0
		160-199	2
		200-239	4
		240-279	5
		>=280	7
	60-69	<160	0
		160-199	1
		200-239	2
		240-279	3
		>=280	4
	70-79	<160	0
		160-199	1
		200-239	1
		240-279	2
		>=280	2
Cigarette smoker	20-39	-	9
	40-49	-	7
	50-59	-	4
	60-69	-	2
	70-79	-	1
All non-smokers	-	-	0
HDL cholesterol (mg/dL)	-	<40	2
		40-49	1
		50-59	0
		>=60	1
Systolic bp (mmHg)	Untreated	<120	0
		120-129	1
		130-139	2
		140-159	3
		>=160	4
Systolic bp (mmHg)	Treated	<120	0
		120-129	3
		130-139	4
		140-159	5
		>=160	6

The algorithm for estimating an individual's 10-year risk for atherosclerotic cardiovascular disease using pooled cohort equation is as follows:

Step 1: Determine the race-sex group of the individual

Step 2: Take the natural log of the risk factors (age, total cholesterol, HDL-C, systolic-untreated, systolic-treated)

Step 3: Multiply each of the values from step 1 by the coefficient of the individual risk factors (Value * Coefficient)

Step 4: Individual sum = Σ (value*coefficient)

Step 5: Compute the probability of 10-year risk

$$1 - S_{10}e^{(Individual\ sum - MeanX'B)} \quad (1)$$

where S_{10} is the baseline survival for a specific race-sex group, and MeanX'B is the product of the mean coefficient and value for a specific race-sex group.

Sandi *et al*. (2016) developed a web-based smart health system for classifying hypertension risk using C4.5 algorithm. The framework which consists of the learning phase, diagnosis phase and prognosis phase; had the diagnosis phase classify the medical parameters into normal, pre-hypertension, hypertension-stage-1 and hypertension-stage-2. The outcome of the diagnosis phase is passed on to the prognosis phase where the final outcome is the health risk of the patient and the corresponding risk factors. The validation of the developed model on 220 instances exhibited an accuracy of 93.6% with a recall of 88.4%, 97.3%, 88.4%, and 83.3% for normal, hypertension-stage-1, hypertension-stage-2, and pre-hypertension, respectively. Whilst recall of hypertension-stage-1 had a good proportion of 97.3%, the misclassification of hypertension-stage-1 patient as normal is highly risky. A classification system that predicts the risk of hypertension is of great significance, but a dreaded ailment like hypertension does require appropriate management through medication and nutrition recommendation. Using

semantic web technology, Clunis (2019) designed personalised recipes for hypertensive patients. With ingredients, preparation instructions and provenance data were web scraped from recipe websites. Ontology 101 was used to create a recipe knowledge-based model for the management of hypertension. The ontology was validated and tested with reasoner and competency questions. The ontology had the ability to answer explicit recipe questions of hypertensive patients. Nevertheless, the management of hypertension without the knowledge of medical and nutritional experts constitutes a great challenge such as consumption of nutrients without hypertension risk evaluation.

Rachata and Temde (2018) constructed a personalised lifestyle recommendation model for patients concurrently suffering from diabetes mellitus and hypertension. The model developed used fuzzy near compactness. The knowledge of patient's condition with their ranges and lifestyle plans were obtained from medical experts, research works and medical handbooks. Trapezoid fuzzy number expressed the membership degree of both patients' conditions and practice plan, and these features were weighted accordingly. The patients' conditions were further mapped to the corresponding practice plans. The model testing was done using data of 121 patients. The recommendation from the model when compare to the recommendation given to those patients by 11 experts, effected 96% accuracy. Similarly, Sookrah *et al.* (2019) built a mobile-based Mauritian diet recommendation system. The system hinges on the fusion of multilayer perceptron classifier and content-based filtering. Depending on the user preference, the model outputs a food category. Content-based filtering recommends meal plan according to the classified food category, DASH serving size and the number of days the user last consumed a particular meal plan. The classifier conferred 99% accuracy. Blood pressure, sodium intake, alcohol and smoking level were used to compute sodium diet plan. However, there are other factors detrimental to hypertensive such as protein intake. Furthermore, the study considers weight as a factor to determine the risk level of a hypertensive condition; But weight does not measure the amount of fat in a human body. According to American Heart Foundation, excess body fat raises blood pressure and blood cholesterol. Moreover, there is a correlation between hypertension and body mass index (BMI), (Duboz *et al*., 2014). BMI provides an easy way of measuring body fat. Nonetheless, medical researchers have debated on the efficiency of BMI in distinguishing muscles from harmful fat. A Plethora of approaches has been suggested but these come with a huge financial burden. Until a more viable means is proven, BMI remains the best option.

Hui and Sufahani (2019) used both linear programming and integer programming to develop a mathematical model that recommends a low-cost diet that meets the requirement of hypertension patients. Linear programming offered low cost of food items on the number of dishes slightly above the recommended number of dishes while integer programming generated menu with the required number of dishes and a slight increase in cost. The study considered age, gender, physical activity level, health stage, food price, nutrient composition of each food item and recommended nutrient intake for Malaysians as features. However, the medical features considered cannot solely determine the risk level of the blood pressure.

From the foregoing literature, some studies applied content-based filtering algorithm in proposing a menu recommendation system based on the preference of the user. The ability of recommendation from a content-based recommendation system is limited, if the system does not contain enough information on the likes and dislikes of the user (Pazzani and Billsus, 2007). Besides meeting the nutritional preference of the patient which may not necessarily be healthy, it is pertinent to recommend diets that conform with the optimal food for hypertensive patients, taking into consideration the patient's absolute risk. In otherwords, the prevention intensity should match the individual's overall risk. Unlike other domain recommendation systems, diet recommendation system should be attentive to other factors in addition to users' preference such as prior diet consumption (Zenun 2017). When considering a patient's diet history, the process of inferring a nutrition advice is not discrete but consists of uncertainties (Lee *et al*., 2014).

MATERIALS AND METHODS

Data collection Procedure

The target population proposed for enrolment into this study are adult's residents or patients(> 18 years) from the rural and urban areas of Akwa Ibom State, Nigeria–the study area. Participants may be enrolled through primary health centres, hospitals, and the university campuses. WHO stepwise approach surveillance–ageneric and standardised questionnaire-based tool for collecting and measuring non-communicable disease risk is adopted for data gathering. Features considered for this study include demographic, anthropometric, biochemical, and behavioural. Demographic features comprise of age, gender, income level and the highest level of education attained. For every participant, determining obesity encompasses two steps. First step is the measuring of the waist circumference (WC) equidistance between crest of ilium and coastal arch. Joint Scientific statement on Harmonising the Metabolic Syndrome WC threshold for SubSaharan Africa is used to categorise the various waist circumference measurements; men with WC > 94 while women with WC >80 are considered obese. The final step calculates the body mass index (BMI) from the height and weight of the participant using stadiometer and a beam balance, respectively. The BMI is classified according to the standard of Centre for Disease Control and Prevention as follows: underweight (BMI < 18.5 kg/m^2), normal weight (18.5 kg/m^2>= BMI < 25.0 kg/m^2), overweight (25.0 kg/m^2>= BMI <30 kg/m^2), obese (BMI > 30.0 kg/m^2). Moreover, obese BMI is further grouped into 3 classes; class 1 (30.0 kg/m^2>= BMI < 35.0 kg/m^2), class 2 (35.0 kg/m^2 >= BMI < 40 kg/m^2), class 3 is considered extreme obesity (BMI >=40.0 kg/m^2). For heart rate and blood pressure readings, the average of 3systolic and diastolic blood pressure readings obtained from participants in a relaxed and seated position using sphygmomanometer with an appropriate cuff size. Highlighted in the 2017 Hypertension Clinical Practice Guidelines, a wrong cuff size may result in either artificially elevated or lower BP readings if the cuff size is too small or too large. Assessing lipid profile parameters (total cholesterol), a venepuncture of blood sample through the finger will be obtained and analysed on cholesterol analyser. Classification of the lipid profile parameters will be based on the guidelines established by International Diabetes Federation; hypercholesterolemia (values >= 200 mg/dL), hypertriglyceridemia (values >= 150 mg/dL), values<= 40mg/dL indicates low HDL in men and values <= 50mg/dL specifies low HDL in women. Finally, behavioural measurement is obtained including information on tobacco consumption, alcohol consumption, diet, and physical activity level.

Features will be extracted from the raw data obtained through questionnaire. Redundancy and outliers will be eliminated at this level and the data presented in a format readable by machine learning algorithm. To avoid over-fitting due to too many features, dimension reduction and normalisation techniques will be applied to streamline and select the minimal number of features that will discover existing and vital patterns in the data.

Patient Cohort Simulation

In this section, we perform a Monte Carlo simulation of patient cohorts from the Framingham's criteria. We choose Framingham because the pooled cohort equation lacks variable characterisation and scores, to derive complete data. We propose a reverse engineering approach that derives the cohort data from criteria established from a study conducted on patients with/without CVD. The various columns (BCOL to LCOL) represent the risk factors and Framingham's scores for the various CVD. The MCOL is the total Framingham's score. The simulation was done in Microsoft Excel using the uniform probability distribution function as the random number generator. The Excel functions for computing the individual columns are given in Fig. 1.

```
COMPUTE AGE RANGE(BCOL): {=RANDBETWEEN(20,79)}
COMPUTE POINT FOR AGE (CCOL):
    {=IF(B2>=75,16,IF(B2>=70,14,IF(B2>=65,12,IF(B2>=60,10,IF(B2>=55,8,IF(B2>=50,6,IF(B
    2>=45,3,IF(B2>=40,0,IF(B2>=35,-3,IF(B2>=20,-7,0)))))))))}
COMPUTE TOTAL COLESTEROL(DCOL):
    {=IF(B2>=70,IFS(E2=0,RANDBETWEEN(40,159),E2=1,RANDBETWEEN(160,239),E2=2,RANDBETWEE
    N(240,300)),IF(B2>=60,IFS(E2=0,RANDBETWEEN(40,159),E2=1,RANDBETWEEN(160,199),E2=2,
    RANDBETWEEN(200,239),E2=3,RANDBETWEEN(240,279),E2=4,RANDBETWEEN(280,300)),IF(B2>=5
    0,IFS(E2=0,RANDBETWEEN(40,159),E2=2,RANDBETWEEN(160,199),E2=4,RANDBETWEEN(200,239)
    ,E2=5,RANDBETWEEN(240,279),E2=7,RANDBETWEEN(280,300)),IF(B2>=40,IFS(E2=0,RANDBETWE
    EN(40,159),E2=3,RANDBETWEEN(160,199),E2=6,RANDBETWEEN(200,239),E2=8,RANDBETWEEN(24
    0,279),E2=10,RANDBETWEEN(280,300)),IF(B2>=20,IFS(E2=0,RANDBETWEEN(40,159),E2=4,RAN
    DBETWEEN(160,199),E2=8,RANDBETWEEN(200,239),E2=11,RANDBETWEEN(240,279),E2=13,RANDB
    ETWEEN(280,300)),0)))))}
COMPUTE POINT FOR TOTAL COLESTEROL(ECOL):
    {=IF(B2>=70,CHOOSE(RANDBETWEEN(1,3),0,1,2),IF(B2>=60,CHOOSE(RANDBETWEEN(1,5),0,1,2
    ,3,4),IF(B2>=50,CHOOSE(RANDBETWEEN(1,5),0,2,4,5,7),IF(B2>=40,CHOOSE(RANDBETWEEN(1,
    5),0,3,6,8,10),IF(B2>=20,CHOOSE(RANDBETWEEN(1,5),0,4,8,11,13),0)))))}
COMPUTE SMOKE STATUS (0-NO, 1-YES) (FCOL): {=RANDBETWEEN(0,1)}
COMPUTE POINT FOR SMOKE STATUS (GCOL):
    {=IF(F2=1,(IF(B2>=70,1,IF(B2>=60,2,IF(B2>=50,4,IF(B2>=40,7,IF(B2>=20,9,0)))))),0)}
COMPUTE HDL COLESTEROL (HCOL): {=RANDBETWEEN(30,70)}
COMPUTE POINT FOR HDL COLESTEROL (ICOL):
    {=IF(H2>=60,1,IF(H2>=50,0,IF(H2>=40,1,IF(B2>=0,2,0))))}
COMPUTE SYSTOLIC BP STATUS (1-TREAT, 0-UNTREAT) (JCOL): {=RANDBETWEEN(0,1)}
COMPUTE SYSTOLIC RANGE (KCOL): {=RANDBETWEEN(80,210)}
COMPUTE POINT FOR SYSTOLIC RANGE (LCOL):
    {=IF(J2=1,(IF(K2>=160,6,IF(K2>=140,5,IF(K2>=130,4,IF(K2>=120,3,IF(K2>=0,0)))))),IF
    (J2=0,(IF(K2>=160,4,IF(K2>=140,3,IF(K2>=130,2,IF(K2>=120,1,IF(K2>=0,0)))))))))}
COMPUTE TOTAL POINT (MCOL): {=C2+E2+G2+I2+L2}
```

Fig. 1. Excel functions for simulating Framingham scores or variables on Tables 1 and 2

Patient Cohort Clustering

Self-Organising Maps (SOMs)–an unsupervised artificial neural network (ANN), (Kohonen, 1982) have been applied to several problems. This simple yet powerful algorithm can reduce incredibly complex problems down to easily interpreted data mappings. The main drawback of the SOM is that it requires that the neuron weights be necessary and sufficient to cluster inputs. The SOM learning algorithm is relatively straightforward. It consists of initialising the weights, iterating over the input data, finding the "winning" neuron for each input, and adjusting the weights based on the location of that "winning" neuron. During mapping, there will be one single winning neuron: the neuron whose weight vector lies closest to the input vector. This can be simply determined by calculating the Euclidean distance between input vector and weight vector. A pseudo code implementing the SOM is presented in the following steps:

1. *Initialise weights*
   For 0 to X number of training epochs
      Select a sample from the input data set
      Find the "winning" neuron for the sample input
      Adjust the weights of nearby neurons
   End for loop
2. *Find the "Winning" Neuron*
   The algorithm used to find the "winning" neuron is the Euclidean distance equation. That is, the "winning" neuron n of dimension d for sample input v (also of dimension d) would be the one which minimised the equation:

$$\omega = \sqrt{\sum_{i=0}^{d}(n_i - v_i)^2} \tag{2}$$

RESULTS

Patients Cohort Simulation

A simulation run of 10,000 patients was performed. Sample computations of the first 20 simulations for male and female patients are presented in Tables3 and 4, respectively.

Table 3. Result of male cohorts' simulation

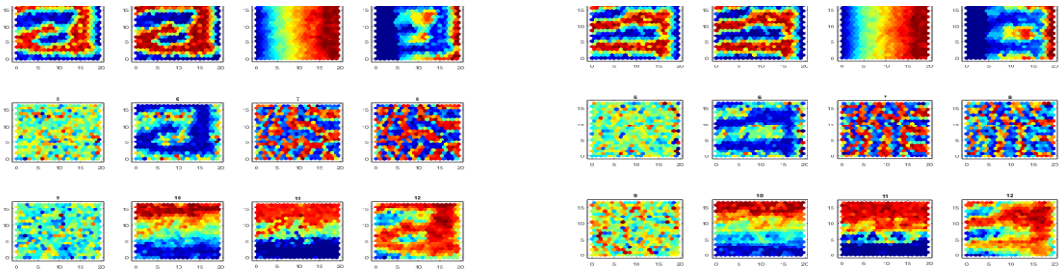
A	B	C	D	E	F	G	H	I	J	K	L	M	N
Patient	Age	F-Scr	Total Chol.	F-Scr	Smoke? [0-no, 1-yes]	F-Scr	HDL Col.	F-Scr	Sys. Bp? [1-Treat, 0-Untreat]	Sys Range	F-Scr	Total F-Scr	%
1	50	6	98	0	0	0	61	1	1	152	5	12	1.00
2	64	10	256	3	1	2	62	1	1	196	6	22	17.00
3	55	8	226	4	1	4	34	2	0	206	4	22	17.00
4	64	10	211	2	0	0	32	2	1	188	6	20	11.00
5	45	3	84	0	0	0	70	1	0	210	4	8	0.46
6	79	16	264	2	0	0	44	1	1	201	6	25	78.00
7	37	-3	176	4	0	0	70	1	1	148	5	7	0.06
8	59	8	293	7	1	4	49	1	1	141	5	25	72.00
9	62	10	214	2	1	2	56	0	1	157	5	19	8.00
10	75	16	173	1	0	0	41	1	1	92	0	18	6.00
11	27	-7	284	13	1	9	31	2	0	132	2	19	8.00
12	27	-7	283	13	1	9	45	1	0	190	4	20	11.00
13	67	12	195	1	0	0	32	2	1	210	6	21	14.00
14	38	-3	262	11	1	9	33	2	1	82	0	19	8.00
15	73	14	238	1	0	0	52	0	0	204	4	19	8.00
16	72	14	58	0	0	0	68	1	1	130	4	19	8.00
17	76	16	50	0	0	0	31	2	0	122	1	19	8.00
18	45	3	273	8	0	0	36	2	1	101	0	13	2.00
19	70	14	249	2	0	0	70	1	1	192	6	23	22.00
20	42	0	300	10	1	7	34	2	0	125	1	20	11.00

Table 4. Result of men cohorts' simulation

A	B	C	D	E	F	G	H	I	J	K	L	M	N
Patient	Age	F-Scr	Total Chol.	F-Scr	Smoke? [0-no, 1-yes]	F-Scr	HDL Col.	F-Scr	Sys. Bp? [1-Treat, 0-Untreat]	Sys Range	F-Scr	Total F-Scr	%
1	49	3	142	0	1	5	40	1	0	179	2	11	1.00
2	35	-4	225	7	1	8	66	-1	0	115	0	10	1.00
3	33	-9	169	4	0	0	61	-1	1	178	3	-3	0.28
4	26	-9	284	11	0	0	68	-1	0	81	0	1	0.81
5	23	-9	129	0	0	0	39	2	1	180	3	-4	0.37
6	32	-9	196	4	1	8	33	2	1	140	2	7	0.21
7	51	6	235	3	0	0	62	-1	0	119	0	8	0.32
8	57	8	268	4	1	3	41	1	1	154	2	18	6.00
9	43	0	171	3	0	0	68	-1	1	108	0	2	0.51
10	41	0	238	5	0	0	43	1	0	177	2	8	0.12
11	49	3	88	0	1	5	35	2	1	105	0	10	1.00
12	25	-9	270	9	1	8	46	1	1	95	0	9	1.00
13	39	-4	170	4	1	8	54	0	0	148	1	9	1.00
14	29	-9	218	7	1	8	63	-1	1	115	0	5	0.53
15	29	-9	282	11	1	8	37	2	0	166	2	14	2.00
16	24	-9	176	4	1	8	55	0	0	147	1	4	0.15
17	30	-9	246	9	0	0	35	2	1	123	1	3	0.44
18	70	12	234	0	0	0	37	2	0	194	2	16	4.00
19	44	0	172	3	1	5	51	0	0	162	2	10	1.00
20	67	11	290	3	0	0	32	2	0	172	2	18	6.00

Patient Cohort Pattern Visualisation

Fig. 2 shows SOM component planes results generated from the simulated data, using a batch unsupervised ANN learning model in MATLAB 2015b environment. Each component reveals pattern exhibited by the simulated data for male cohorts (Fig 2a) and female cohorts (Fig. 2b). Observe that age, Framingham's age score, Framingham's smoking score and total Framingham's score exhibit similar patterns but with varying cluster boundaries. As expected, smoking status and systolic BP status exhibit similar patterns. HDL-cholesterol and its Framingham's score also exhibit similar patterns for male and female patients; and same goes for systolic BP and its Framingham's score.



(a) Male patients (b) Female patients

Fig. 2. SOM component planes for patient cohorts. 1-Age; 2-Framingham's score for age; 3-Total cholesterol; 4-Framingham's score for total cholesterol; 5-Smoking status; 6- Framingham's score for smoking status; 7-HDL cholesterol; 8-Framingham's score for HDL cholesterol; 9-Systolic BP status; 10-Systolic BP; 11- Framingham's score for systolic BP; 12-Total Framingham score.

Risk Factors Component Planes Analysis

To study the degree of association between the risk factors, we decoupled the SOM correlation hunting matrix into two matrices representing male and female cohorts on Table 5 and Table 6, respectively. Observe that age and systolic BP strongly correlate with their Framingham's scores (i.e., age: r^2=0.97 for male patients and r^2=0.99 for female patients; systolic BP: r^2=0.88 for male patients and r^2=0.87 for female patients).

Table 5. Risk-factor correlations for male patients

	COLB	*COLC*	*COLD*	*COLE*	*COLF*	*COLG*	*COLH*	*COLI*	*COLJ*	*COLK*	*COLL*	*COLM*
COLB	1.00											
COLC	0.97	1.00										
COLD	0.01	-0.01	1.00									
COLE	-0.63	-0.63	0.63	1.00								
COLF	0.01	-0.01	0.00	-0.01	1.00							
COLG	-0.46	-0.48	0.00	0.32	0.72	1.00						
COLH	-0.01	0.02	-0.01	0.00	-0.02	0.02	1.00					
COLI	-0.01	0.01	0.02	0.00	-0.02	0.01	-0.97	1.00				
COLJ	0.00	0.01	0.00	-0.01	0.01	-0.01	0.00	0.01	1.00			
COLK	0.01	0.00	0.00	0.00	0.01	-0.02	-0.01	-0.01	0.00	1.00		
COLL	-0.02	-0.01	0.01	0.00	0.00	0.01	0.00	-0.01	0.30	0.88	1.00	
COLM	0.73	0.76	0.35	-0.13	0.36	0.01	-0.19	0.18	0.04	0.16	0.17	1.00

Table 6. Risk-factor correlations for female patients

	COLB	*COLC*	*COLD*	*COLE*	*COLF*	*COLG*	*COLH*	*COLI*	*COLJ*	*COLK*	*COLL*	*COLM*
COLB	1.00											
COLC	0.99	1.00										
COLD	-0.07	-0.07	1.00									
COLE	-0.57	-0.58	0.72	1.00								
COLF	0.01	-0.02	0.00	0.00	1.00							
COLG	-0.43	-0.44	0.03	0.27	0.77	1.00						
COLH	-0.01	0.01	-0.01	-0.01	0.00	0.00	1.00					
COLI	0.00	0.01	0.00	0.02	0.00	0.00	-0.61	1.00				
COLJ	0.00	-0.02	0.00	-0.01	0.00	0.00	-0.01	0.01	1.00			
COLK	0.00	0.00	0.01	0.00	0.00	0.00	-0.01	-0.01	-0.02	1.00		
COLL	-0.03	0.01	0.00	0.00	0.02	0.00	0.00	0.00	0.30	0.87	1.00	
COLM	0.62	0.63	0.36	0.05	0.39	0.16	-0.06	0.11	0.11	0.28	0.35	1.00

Furthermore, age associates moderately with Framingham's total cholesterol score and total Framingham's score, with correlation coefficient values of -0.63 and 0.73 for male patients, and -0.57 and 0.62 for female patients, respectively. Framingham's score for age also associates moderately with Framingham's total cholesterol score and total Framingham's score, with correlation coefficient values of -0.63 and 0.73 for male patients, and -0.58 and 0.63 for female patients, respectively. Moderate correlations are also observed for total cholesterol and its Framingham' score (r^2=0.63 for male patients and r^2=0.72 for female patients); smoking status and Framingham's smoking score (r^2=-0.72 for male patients and r^2=0.77 for female patients).

A strong negative correlation is observed for HDL-cholesterol and its Framingham's score in male patients (r^2=-0.97) but moderately correlated in female patients (r2=-0.61). This implies that the scoring system is inversely proportional to the factor.

For the discussed associations, negative correlations indicate an opposite direction or magnitude, order, or distribution of the dataset. Such correlations are observed between age and its Framingham's score, total cholesterol and its Framingham's score, and HDL-cholesterol and its Framingham's score, for male and female cohorts.

DISCUSSION

Hypertension has been commonly ascertained as the paramount risk factor for cardiovascular events. Epidemiological data on general population illustrates the effect of increase in age on systolic and diastolic blood pressure. The prevalence of hypertension is thus significantly associated with increase in age, (Anderson, 1999). From the analysis of about 4,800 referred hypertensive patients, Anderson (1999) suggests that systolic blood pressure increases after the age of 60 years. Under laboratory conditions, cigarette smoking has been found to influence the rise in BP, (Green *et al.,* 1986). However, several studies have statistically demonstrated the inverse interaction between smoking and BP, the association between cigarette dosage and BP response has also been studied, (Berglund and Wilhelmsen, 1975; Savdie *et al*., 1984). Lee *et al.* (2001) in a 4-year follow-up study discovered increase in BP among current non-smokers and higher increase among quitters. The rise in BP of quitters of <1 year is quite similar to those of current smokers. However, quitters of >1 year displayed raised BP than current smokers. This observation was consistent across different groups stratified by weight. Hence, cessation of smoking may trigger hypertension (Rupprecht *et al*., 2017). This inference is widely due to the decreased inflammation which may bring about weight gain. Primatesta *et al*. (2001) found a minute effect of smoking on BP. At adjustment of age, BMI, alcohol, and social class, a significantly higher SBP was noticed among older men that are heavy and moderate smokers than non-smokers. Amongst women, light smokers exhibited lower mean BP than heavier smokers and non-smokers. The study inferred no consistent independence in clinical significance of BP among smokers and non-smokers. The observed difference in BP due to smoking is ascribed to age and some confounding factors such as BMI and alcohol. The correlations of risk factors obtained in this study therefore negates what is obtained in the literature as the smoking dataset was found to correlate positively with BP (Table 5 and Table 6). This indicates the inability of the Framingham's measure to model interactions between the various risk factors. Other risk variables such as BMI and alcohol consumption are missing in the Framingham's model, calling for a more comprehensive model.

Ferrara *et al.* (2002) clinically investigated the impact of serum cholesterol on BP levels during ambulatory monitoring, sympathetic stimulation and at rest with no consideration of other well-known determining factors. This investigation unearthed the evidence that a little rise in serum cholesterol influences BP during sympathetic stimulation. Moreover, cholesterol has been pathophysiologically evidence to induce endothelial dysfunction (Hayakawa and Raij, 1999).The Framingham's model also fails to model the interaction between total cholesterol, as evidence in very weak correlations are observed between the two variables. Coca *et al*. (2009) reported the prevalence of low HDL-cholesterol in women with increased risk of cardiovascular event. This interaction between low High-Density Lipoprotein Cholesterol (HDL-C) concentration and CVD was independent of other determinants. The Framingham's model again fails to model the interaction between HDL-C and other CVD factors, as evidence in very weak correlations observed between these variables.

Seemingly, aside age, the Framingham's model relates only the risk factors and their scores, hence, making it an inappropriate model for tackling CVD.

Expert System for Personalised Nutrition Recommendation

We propose in Fig. 3, an expert system/framework for personalised nutrition therapy. Our expert system is a recommendation system that consists of 4 components namely, Data

Collection/Processing, Knowledge Base, Prediction, Diet Recommendation. We discuss the various components as follows:

Data Collection/Processing–The WHO stepwise approach to non-communicable diseases risk factor surveillance is a questionnaire for gathering risk factor data for all non-communicable diseases. CVD variables can be extracted from this questionnaire to prepare the data collection instrument. We are in the process of collecting and processing clinical data for the building of a CVD database, which will serve as a useful resource for advancing research progress in this area.

Knowledge Base–Measures for input and output linguistic variables are defined by leveraging on the knowledge from literature and human experts. A complete range of values also known as the universe of discourse (UoD) is established and assigned to input and output linguistic variables–a membership function (MF) through fuzzification is used to map each element in a UoD to a membership degree thus quantifying the grade of membership of the element in UoD to a fuzzy set for a particular input variable. The dynamics of modelling interactions between the CVD risk factors are characterised by rules from the expert domain. Fuzzy rule base consisting of IF-THEN logic (linguistic rules) is integral to fuzzy logic systems (FLSs) for simulating the expert reasoning. The rule base of our fuzzy expert system is made up of linguistic rules that map the relationship between the input fuzzy sets defined for CVD risk factors and output linguistic variables, consequently, inferring the resultant CVD risk response at the fuzzy inference engine. The defuzzification interface of our FLS uses the membership function to establish the CVD risk (crisp value) from the output linguist variables (CVD risk response).

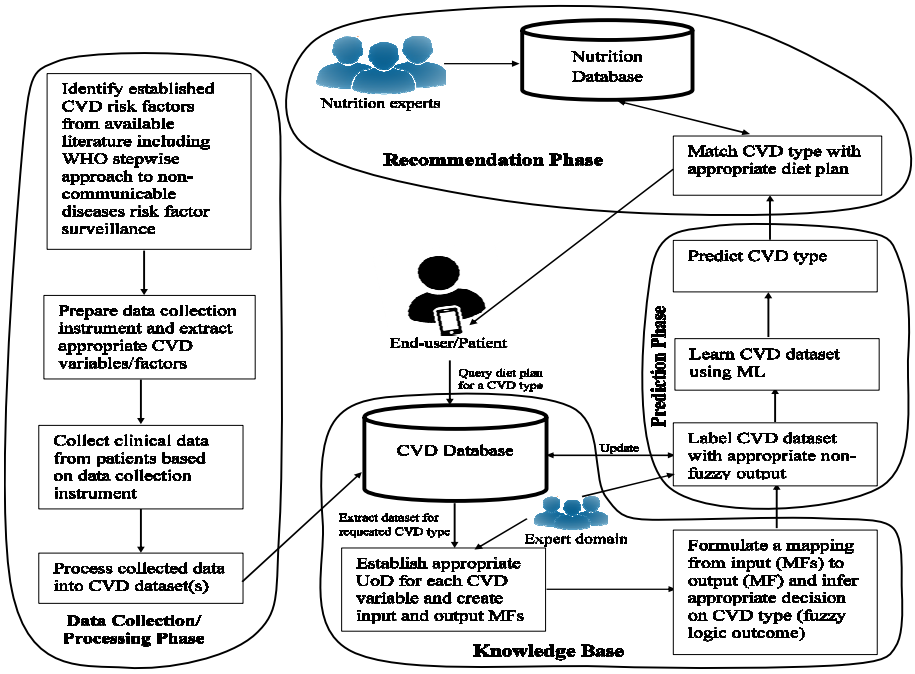


Fig. 3. Proposed expert recommendation system framework

CVD Prediction–Prior to training, validating, and testing of the machine learning algorithm, the CVD risk factor dataset is updated with the achieved CVD risk level or type. At this stage, the knowledge of medical experts is employed in the stratification and labelling of the risk classes as (very low risk, low risk, moderate risk, high risk and very high risk). The updated CVD risk factor dataset is then partitioned into train-validate-test dataset. The training-validate dataset is used to iteratively learn and validate the classification algorithm. For optimal performance, during validation, evaluation metrics are used to assess the performance of the model. The test

dataset is finally deployed to test the performance of the classification algorithm, to discover how generic the model is, and to predict the CVD risk class.

Recommendation–The risk class is matched with appropriate diet plan in the nutrition database crafted by nutrition experts or dieticians, for diet recommendation that will serve as a therapy for curbing or mitigating the chance of developing CVD.

Conclusions and Future Research Direction

Prognosticating the risk of hypertension in seemingly healthy/unhealthy individuals and prescribing nutritional therapy to attenuate the latent health hazard due to hypertension is a momentous research. The robustness of CVD risk predictor is highly dependent on the sample data used in developing the algorithm. On this basis, we assessed the variable interaction of the data cohorts generated using Framingham's algorithm and discover no interaction among other CVD variables, save age. We proposed an integrative procedure consisting of fuzzy inference system, to create the necessary interaction between CVD variables and machine learning classifier, for precise prediction of risk level.

A future research direction is to conduct real cohorts' data collection and implement the proposed expert recommendation system. The deployed system will certainly impact the health policy and serve as meaningful decision support to CVD risk level prediction and diet plan recommendations, for dieticians and people living with CVD.

ACKNOWLEDGEMENT

This research is funded by the Tertiary Education Trust Fund (TETFund), under the Institutional Based Research (IBR) Grant.

REFERENCES

Adeloye, D., Basquill, C., Aderemi, A. V., Thompson, J. Y., and Obi, F. A. (2015). An estimate of the prevalence of hypertension in Nigeria: a systematic review and meta-analysis. *Journal of hypertension*, *33*(2), 230-242.

Anderson, T. J., Grégoire, J., Pearson, G. J., Barry, A. R., Couture, P., Dawes, M., and Hegele, R. A. (2016). 2016 Canadian Cardiovascular Society guidelines for the management of dyslipidemia for the prevention of cardiovascular disease in the adult. *Canadian Journal of Cardiology*, *32*(11), 1263-1282.

Berglund, G., and Wilhelmsen, L. (1975). Factors related to blood pressure in a general population sample of Swedish men. *Acta medica scandinavica*, *198*(1□6), 291-298.

Boateng, D., Agyemang, C., Beune, E., Meeks, K., Smeeth, L., Schulze, M. B., ... and Danquah, I. (2018). Cardiovascular disease risk prediction in sub-Saharan African populations—Comparative analysis of risk algorithms in the RODAM study. *International journal of cardiology*, *254*, 310-315.

Clunis, J. (2019). Designing an ontology for managing the diets of hypertensive individuals. *International Journal on Digital Libraries*, *20*(3), 269-284.

Delong C. and Sharma S. (2020). Physiology, Peripheral Vascular Resistance. StatPearls Publishing, PMID: 30855892.https://pubmed.ncbi.nlm.nih.gov/30855892/.

Duboz, P., Boëtsch, G., Gueye, L., and Macia, E. (2014). Hypertension prevalence, awareness, treatment and control in Dakar (Senegal). *Journal of human hypertension*, *28*(8), 489-493.

Egan, B. M. (2018). Defining Hypertension by Blood Pressure 130/80 mm Hg Leads to an Impressive Burden of Hypertension in Young and Middle□Aged Black Adults: *Journal of the American Heart Foundation,*

Gallestery, J. B. (2016). Framingham Heart Study. Encyclopaedia Britannica inc.

Goff, D. C., Lloyd-Jones, D. M., Bennett, G., Coady, S., D'agostino, R. B., Gibbons, R., ... and Robinson, J. G. (2014). 2013 ACC/AHA guideline on the assessment of cardiovascular risk: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. *Journal of the American College of Cardiology*, *63*(25 Part B), 2935-2959.

Green, M. S., Jucha, E., and Luz, Y. (1986). Blood pressure in smokers and nonsmokers: epidemiologic findings. *American heart journal*, *111*(5), 932-940.

Grundy, S. M., Stone, N. J., Bailey, A. L., Beam, C., Birtcher, K. K., Blumenthal, R. S., ... and Goldberg, R. (2018). Guideline on the management of blood cholesterol: a report of the American College of Cardiology/American heart association Task force on clinical practice guidelines. *J Am Coll Cardiol*, *73*(24), e285-350.

Hayakawa, H., and Raij, L. (1999). Relationship between hypercholesterolaemia, endothelial dysfunction and hypertension. *Journal of hypertension*, *17*(5), 611-619.

Hui, L. S., and Sufahani, S. *(2019). Healthy Menu Scheduling for High Blood Pressure Patient with Optimization Method through Integer Programming. Advances in Computing and Intelligent System, 1(1).*

Ko, D. T., Sivaswamy, A., Sud, M., Kotrri, G., Azizi, P., Koh, M., ... and Tu, K. (2020). Calibration and discrimination of the Framingham Risk Score and the Pooled Cohort Equations. *CMAJ*, *192*(17), E442-E449.

Kohonen, Teuvo (1982). Self-Organized Formation of Topologically Correct Feature Maps. Biological Cybernetics. 43 (1): 59–69.

Lee, C. S., Wang, M. H., and Lan, S. T. (2014). Adaptive personalized diet linguistic recommendation mechanism based on type-2 fuzzy sets and genetic fuzzy markup language. *IEEE Transactions on Fuzzy Systems*, *23*(5), 1777-1802.

Lee, D. H., Ha, M. H., Kim, J. R., and Jacobs Jr, D. R. (2001). Effects of smoking cessation on changes in blood pressure and incidence of hypertension: a 4-year follow-up study. *Hypertension*, *37*(2), 194-198.

Luft, F. C. (1998). Molecular genetics of human hypertension. *Journal of hypertension*, *16*(12), 1871-1878.

Moons, K. G., Kengne, A. P., Woodward, M., Royston, P., Vergouwe, Y., Altman, D. G., and Grobbee, D. E. (2012). Risk prediction models: I. Development, internal validation, and assessing the incremental value of a new (bio) marker. *Heart*, *98*(9), 683-690.

Muntner, P., Shimbo, D., Carey, R. M., Charleston, J. B., Gaillard, T., Misra, S., ... and Wright Jr, J. T. (2019). Measurement of blood pressure in humans: a scientific statement from the American Heart Association. *Hypertension*, *73*(5), e35-e66.

Nguyen, H., Odelola, O. A., Rangaswami, J., and Amanullah, A. (2013). A review of nutritional factors in hypertension management. *International journal of hypertension*, *2013*.

Pazzani, M. J., and Billsus, D. (2007). Content-based recommendation systems. In *The adaptive web* (pp. 325-341). Springer, Berlin, Heidelberg.

Preiss, D., and Kristensen, S. L. (2015). The new pooled cohort equations risk calculator. *Canadian journal of Cardiology*, *31*(5), 613-619.

Primatesta, P., Falaschetti, E., Gupta, S., Marmot, M. G., and Poulter, N. R. (2001). Association between smoking and blood pressure: evidence from the health survey for England. *Hypertension*, *37*(2), 187-193.

Pylypchuk, R., Wells, S., Kerr, A., Poppe, K., Riddell, T., Harwood, M., ... and Metcalf, P. (2018). Cardiovascular disease risk prediction equations in 400 000 primary care patients in New Zealand: a derivation and validation study. *The Lancet*, *391*(10133), 1897-1907.

Rachata, N., and Temdee, P. (2018). Personalized Recommendation for Preventing Patients with Type 2 Diabetes Mellitus and Hypertension from Cardiovascular Complication. In *2018 Global Wireless Summit (GWS)* (pp. 294-299). IEEE.

Raji, Y. R., Abiona, T., and Gureje, O. (2017). Awareness of hypertension and its impact on blood pressure control among elderly Nigerians: report from the Ibadan study of aging. *The Pan African Medical Journal*, *27*.

Rupprecht, L. E., Koopmeiners, J. S., Dermody, S. S., Oliver, J. A., Al'Absi, M., Benowitz, N. L., and Donny, E. C. (2017). Reducing nicotine exposure results in weight gain in smokers randomised to very low nicotine content cigarettes. *Tobacco control*, *26*(e1), e43-e48.

Sandi, G., Supangkat, S. H., and Slamet, C. (2016, April). Health risk prediction for treatment of hypertension. In *2016 4th International Conference on Cyber and IT Service Management* (pp. 1-6). IEEE.

Savdie, E., Grosslight, G. M., and Adena, M. A. (1984). Relation of alcohol and cigarette consumption to blood pressure and serum creatinine levels. *Journal of chronic diseases*, *37*(8), 617-623.

Sever, P. S., and Poulter, N. R. (1989). A hypothesis for the pathogenesis of essential hypertension: the initiating factors. *Journal of hypertension. Supplement: official journal of the International Society of Hypertension*, *7*(1), S9-12.

Sookrah, R., Dhowtal, J. D., and Nagowah, S. D. (2019). A DASH Diet Recommendation System for Hypertensive Patients Using Machine Learning. In *2019 7th International Conference on Information and Communication Technology (ICoICT)* (pp. 1-6). IEEE.

Stone, N. J., and Lloyd-Jones, D. M. (2016). The Pooled Cohort equations for predicting risk of myocardial infarction and stroke: validated in representative natural history populations. In *Mayo Clinic Proceedings* (Vol. 91, No. 6, pp. 692-694). Elsevier.

Stone, N. J., Robinson, J. G., Lichtenstein, A. H., Merz, C. N. B., Blum, C. B., Eckel, R. H., ... and McBride, P. (2014). 2013 ACC/AHA guideline on the treatment of blood cholesterol to reduce atherosclerotic cardiovascular risk in adults: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. *Journal of the American College of Cardiology*, *63*(25 Part B), 2889-2934.

Zenun Franco, R. (2017). Online Recommender System for Personalized Nutrition Advice. In *Proceedings of the Eleventh ACM Conference on Recommender Systems* (pp. 411-415).