ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 6, June 2017

Modified Machine Learning Approach for Diabetes Risk Factor Detection

Tejal A. Patil[1], Swati A. Patil[2]

M. E Student, Dept. of Computer Science and Engineering, GHRIEM, Jalgaon Maharashtra, India

Assistant Professor, Computer Science and Engineering, GHRIEM, Jalgaon, Maharashtra, India

ABSTRACT: Diabetes is a deficiency in the body's ability to convert glucose (sugar) to energy. Glucose is the main source of fuel for our body. When food is digested it is changed into fats, protein, or carbohydrates. Foods that affect blood sugars are called carbohydrates. The hypertriglyceridemic waist (HW) is strongly associated with type 2 diabetes Phenotype; however, to date, no study has assessed the predictive power of phenotypes based on individual triglyceride and anthropometric measurements. The aims of the study were to assess the association involving the HW phenotype and type 2 diabetes in Korean adults and to evaluate the predictive power of dissimilar phenotypes consisting of combinations of individual anthropometric measurements and Triglyceride levels. Study measured fasting plasma glucose and TG levels and performed anthropometric measurements. We employed binary logistic regression (LR) to examine statistically significant difference between normal subjects and those with type 2 diabetes using Hypertriglyceridemic waist and individual anthropometric measurements. For more reliable prediction results, two machine learning algorithms, naive Bayes and LR, were used to evaluate the predictive power of various phenotypes.

KEYWORDS: Data Mining, Anthropometric measurements, Phenotype, Type 1 Diabetes Type 2 Diabetes, Naive Bayes, C4.5

I. INTRODUCTION

Data mining is a key role in the intelligent health domain [1]. There are several software's and tools have been used to diagnose and classifies health information's based on the attributes. The huge size databases are included into this process as input. The followings are the Basic information's about the diabetes and its basic causes and symptoms. Diabetes risk Prediction Model can support medical professionals and practitioners in predicting risk status based on the clinical data records. In biomedical field data mining and its techniques plays an essential role for prediction and analysing different type of health issues. The healthcare industry gives huge amounts of healthcare data and that need to be mined to ascertain hidden information for valuable decision selection. Determining hidden patterns and relationships may often very tough and unreliable. The health record is classified and predicted if they have the symptoms of Diabetes risk and using risk factors of disease [2]. It is indispensable to find the best fit algorithm that has greater accuracy, speedy and memory utilization on prediction in the case of Diabetes.

DIABETES:
Diabetes is classified into two types:
1.1 Type 1 Diabetes:

It is a chronic condition in which the pancreas produces little or no insulin. This type of Diabetes results from the pancreas's failure to produce enough insulin. This necessitates the individual to insert insulin or carry an insulin pump. This form was previously referred to as "insulin-dependent diabetes mellitus"(IDDM).

It's uncommon. Only about 5% of people with diabetes have type 1. It's more ordinary in whites than in African-Americans. It affects men and women equally. Although the disease usually starts in people under 20, it can happen at any age.

1.2 Type 2 Diabetes:

Begin with insulin resistance, a condition in which cells be unsuccessful to respond to insulin properly. As the disease progresses a lack of insulin may also develop. This construction was previous to referred to as "non insulin-dependent diabetes mellitus" or "adult-onset diabetes". The primary cause is excessive body Weight and not enough exercise.

It begins with insulin resistance, a condition in which cells be unsuccessful to respond to insulin properly. As the disease progresses a lack of insulin may also develop. This organization was before referred to as "non insulin-dependent diabetes mellitus" or "adult-onset diabetes". The primary cause is excessive body

II. LITERATURE SURVEY

2.1 Prediction of Fasting Plasma Glucose Status Using Anthropometric Measures for Diagnosing Type 2 Diabetes [7]

It is known that body fat distribution and obesity are significant risk factors for type 2 diabetes. Prediction of type 2 diabetes using a grouping of anthropometric measures remains a controversial subject. This study aims to calculate the fasting plasma glucose status that is used in the analysis of type 2 diabetes by a combination of various measures among Korean adults.[2] A total of 4872 subjects (2956 females and 1916 males) participated in this study. Based on 38 anthropometric measures, we compared predictions of FPG status using individual against joint measures using two machine learning algorithms. The standards of the area under operating feature curve in the predictions by logistic regression and naive Bayes classifier based on the grouping of measures were 0.742 and 0.738 in females, respectively, and were 0.686 and 0.685 in males, respectively. Our results indicate that prediction of FPG status using a grouping of anthropometric measures was better to individual measures alone in both females andmales.We show that using balanced data of normal and high FPG groups can progress the prediction and reduce the intrinsic bias of the model toward the majority class. Subjects In this study, we analyzed data from 4872 subjects -(2956 females and 1916 males) aged 31–90 years from the Korean Health and Genome Epidemiology study database (KHGES). The KHGES routinely measures weight, height, and circumferences of the regional sites of the body by trained observers according to standardized protocols. Weight is measured to the nearest 0.1 kg using digital scales, height is measured to the nearest 0.1 cm in barefoot and wearing slight clothes only, and circumferences of nine regional sites of the body from forehead to hip are measured to the nearest 0.1 cm in the standing position using a meter tape. BMI is calculated as the weight (kg) divided by the square of the height (m). The circumference of eight local sites of the body are calculated at the levels of the glabella and occiput (forehead circumference, FC), the thyroid cartilage and cricoid cartilage (neck circumference, NC), the left and right axilla (axillary circumference, AC), the left and right nipples (chest circumference, CC), the left and right seventh and eighth prominence of costochondral junction (rib circumference, RC), the umbilicus (WC), the left and right anterior superior iliac spines (pelvic circumference, PC), and the upper edge of the pubis (hip circumference, HC) [1]. After measuring the circumferences of the eight regional sites, we calculated the ratio between the two sites and obtained a total of 25 ratios.

To diagnose type 2 diabetes, we used the recommendations of the 1990 World Health Organization report [2] and the American Association of Clinical Endocrinologists [4]. Type 2 diabetes is defined as FPG of larger than 110 mg/dl (high FPG status) and/or physician-diagnosed. We carried out all statistical analyses and categorization experiments for females and Males individually, because gender is a significant effect modifier in the association of anthropometric measures with incident diabetes [5]. Baseline characteristics by measures and gender. National Institute of Health Ethics and the Institutional Review Board of the Korean Health. The study and all the subjects gave written informed consent.

2.2 The Best Central Adiposity Index in the Prediction of Cardiovascular Risk Factors [3]

To determine the best index of central obesity those predict cardiovascular risk factors Methods: A cross sectional study involving 918(444 males and 474 females) participants of a community health survey in Sagamu and Remo North Local Government Areas of Ogun State, Nigeria. The body mass index, waist circumference, waist to hip ratio and waist to height ratio (WHR) of the participants are determined by standard protocols. Pearson association between

BMI and the three central obesity index was determined. The area under the curve (AUC) on the ROC was used to determine the best measure of central obesity which identified individuals with general hypertension and obesity. Results: WHtR and WC were better than WHR at detecting the presence of both general obesity and hypertension in both males and females.

2.3 BMI is strongly associated with Waist Circumference and hypertension is strongly associated with type 2 diabetes and dyslipidemia, in northern Chinese adults [4]

To estimate the changes body mass index and waist circumference and their relations with the occurrence of hypertension and type 2 diabetes mellitus (T2DM). Design 2 consecutive population based cross sectional surveys. Setting A total of 14 districts and seven counties in Shanghai, China. Participants 12 320 randomly selected participants of the survey in 2003–2004, and 7423 randomly selected participants of the survey in 2009. All participant be populace of Shanghai aged 35–74 years.

2.4 Waist to height ratio is the best indicator for undiagnosed type 2 diabetes mellitus [1]

Early detection of diabetes is significant for the prevention of diabetic complications. The best obesity index for indicating Type 2 diabetes mellitus remains unclear. We expected to classify the optimal adiposity calculate among BMI, waist circumference, waist-hip ratio and waist to height ratio to specify undiagnosed Type 2 diabetes and impaired fasting glucose in Chinese adults. A total of 7568 participants aged 20-79 years were included in this study. The fasting glucose was defined as a fasting plasma glucose level of 6.1-6.8 mmol/l in participants without diabetes. Undiagnosed Type 2 diabetes was identified as fasting plasma glucose ≥ 7.0 mmol/l when neither a history of diabetes nor use of hypoglycemic drugs was present. Body weight, height, waist and hip circumferences were measured following standard procedures. Data were analyzed using logistic regression and areas under the receiver operating characteristic curves.

2.5 Identifying obesity indicators which best correlate with type 2 diabetes mellitus. [2]

Obesity has exposed to be a analytical indicator of type 2 diabetes (T2D); at rest, the power of different obesity indicators in the detection of type 2 diabetes (T2D) remains controversial. This study evaluates the detecting power of the body mass index (BMI), waist circumference (WC), waist to hip ratio (WHR) and waist to height ratio (WHTR) for the presence of type 2 diabetes (T2D) in undiagnosed diabetics among the Chinese population. Individuals were selected from ongoing large-scale population based Beijing Community Pre Diabetes (BCPD) study cohort. The oral glucose tolerance tests (OGTT) are performed to diagnose diabetes. A total of 230 new cases of T2D and 1,868 normal blood glucose subjects were analyzed

2.6 Waist Circumference Rather than Body Mass Index is Better Indicator of Insulin Resistance in Type 2 Diabetes. [5]

Obesity and insulin resistance are associated with type 2 diabetes mellitus. Obesity can be quantified by body mass index (BMI) and waist circumference (WC). In the same way, insulin resistance (IR) is commonly quantified by the fasting plasma insulin(FPI) and Homeostatic model assessment (HOMA-IR). We expected our study to find correlation between obesity parameters and insulin resistance (IR) especially in the Indian population where even with lower BMI there is more prevalence of type2 diabetes mellitus. In 35 uncomplicated patients of type 2 diabetes mellitus weight and WC were measured and BMI was calculated. Homeostatic model assessment (HOMA-IR) and fasting-plasma-insulin (FPI) level were estimated to assess IR. Significant relationship was found between HOMA-IR and WC but it was non-significant between BMI and HOMA-IR. Correlations also not significant between WC and FPI or BMI. In conclusion, HOMA-IR and WC are superior measures of IR and obesity as compared to FPI and BMI, respectively in type 2 diabetes mellitus. This was a cross-sectional study. The study set of rules was approved by Institutional Ethics Committee, All India Institute of Medical Sciences; New Delhi. Consecutive male patients with clinically apparent type 2 diabetes mellitus. The diagnosis of type 2 diabetes mellitus was famous as per the criteria of American Diabetes

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 6, June 2017

Association (ADA). Subjects with hypertension, ischemic heart disease, stroke, any acute or constant respiratory disorder and tobacco users were excluded. Subjects were of within the age range of 35-65 years.

2.7 Other risk factors and Hypertension associated with cardiovascular diseases among adults [6]

Objective: to identify the prevalence of hypertension and its connection with cardiovascular risk factors among adults. Method: Population Based, cross sectional, descriptive study conducted with 408 adult persons. Data were collected through a survey and measurements of weight, height and waist circumference. Person's Chi-square and multiple logistic regressions were used in the data analysis. Results: 23.04% of the individuals reported hypertension with a higher prevalence among women. Probability Ratio indicated that smoking, body mass index (BMI), waist circumference, diabetes mellitus and dyslipidemia were positively associated with hypertension. Conclusion: high character report hypertension and its association with additional cardiovascular risk factors such as diabetes, obesity and dyslipidemia show the need for specific nursing intervention and the execution of protocols focused on minimizing complications arising from hypertension, as well as to prevent the emergence of other cardiovascular diseases This Population Based, cross sectional, descriptive study was conducted with adult individuals living in Paiçandu, PR, Brazil. This town has a total area of 170.65 km and an estimated population of 35,942 inhabitants, 19,775 of which are adults aged between 20 and 59 years old.

2.8 The association of hypertriglyceridemic waist (HW) phenotype with type 2 diabetes mellitus (T2DM) among individuals with first relative history of diabetes.

Background: Anthropometric measures with biochemical indicators used as screening tools for metabolic abnormalities in adolescents and adults. A only some studies have assessed the relation of Enlarge waist Elevated triglyceride (EWET) diabetes with Phenotype, especially among individuals with first relation history of diabetes. This study expected to estimate the connection of EWET phenotype with diabetes among individuals with history of diabetes. Anthropometric and biochemical measurements were evaluated in a population based cross sectional study of 332 male and 991 female Isfahani adults aged 35-55 year. The extend waist Elevated triglyceride phenotype is define as serum trigcylglycerol concentrations ≥150 mg/dl and concurrent waist circumference (WC) ≥88 cm in females and ≥102 cm in males.

III. PROPOSED ALGORITHM

In Proposed System we identified The Type 1 and Type 2 Diabetes Risk Factor Detection for that we used in this study C 4.5 and Backprapogation Algorithms. In existing System Naïve Bayes and logistic regression are used.

3.1 Proposed Architecture

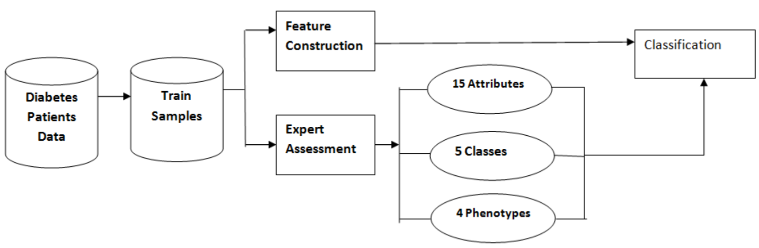

Fig1: System Architecture

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 6, June 2017

Fig 1 shows the System architecture of Type 2 and Type 1 Diabetes Risk Factor Detection.

In the System architecture first we take the type 1 Diabetes Patient Data then Train that sample using naïve Bayes and C4.5 Classifier in that case the 15 Attributes are use these are Age, Sex, HBA1C, Resting B.P, Plasma Glucose, Cholesterol, Date, Pulse Rate, Hypertension, Heredity, Foot Ulcer, Severity index, Phenotypes, Treatment. And 5 Classes these are phenotypes Auxology, Osteometry, Anthropometry, Triglycerides and undetection for Type 1 Diabetes and for Type 2 Diabetes the classes are Interferon, Cytotoxic, Chemokine, and Undetection.

Constructing good features from EHR is often a must to warranty good prediction performance either for expert algorithms or machine learning-based models. This is because raw EHR data are often noisy, sparse, and contain unstructured information that is not directly "computable". Traditional researches on identifying subjects with and without T2DM were using selection strategies built on three features: diabetic diagnosis, diabetic laboratory tests and diabetic medications extracted from EHRs of investigated samples [1][6]. Such researches are limited due to their high missing rates on identification of cases and controls. This is because such strategies applied a conservative selection criterion on cases and controls and were tested in a broader separation range between cases and controls.

3.2 Feature summarization

These features are highly correlated with each other, which will influence performances of computational models to do classification [1][7][2]. And thus we merge correlated features into one feature by summarizing them. Which represents the total number of times T1DM diagnosis and T2DM diagnosis. By using the same way, we summarize all similar features across the seven sources into features. At the same time, the features within a source are also correlated.

3.3 Classification

We use several widely-used classification models such as Naïve Bayes (NB), C4.5and Back Propagation to model patterns of cases and controls based on our features and then use the models to test the ability of our extracted features on identifications of T1DM and T2DM subjects. These classification models are frequently utilized in a wide range of fields, and are recognized as popular choices for classification tasks [2][2][7].

3.4 Naïve Bayes and C4.5 Classifier:-

3.4.1 The Naive Bayesian Classifier:

1. Let T be a training set of samples, each with their class labels. There are k classes, C_1, $C_2 \ldots C_k$. Each sample is represented by an n-dimensional vector, $X = \{x_1, x_2. \ . \ . x_n\}$, depicting n measured values of the n attributes, A_1, A_2, . . . , A_n, respectively.
2. Given a sample X, the classifier will predict that X belongs to the class having the highest a posteriori probability, conditioned on X. That is X is predicted to belong to the class C_i if and only if

$$P(C_i|\mathbf{X}) > P(C_j|\mathbf{X}) \qquad \text{for } 1 \le j \le m, \ j \ne i.$$

Thus we find the class that maximizes P $(C_i|X)$. The class C_i for which P $(C_i|X)$ is maximized is called the maximum posteriori hypothesis. By Bayes' theorem

$$P(C_i|\mathbf{X}) = \frac{P(\mathbf{X}|C_i)\ P(C_i)}{P(\mathbf{X})}.$$

3. As P (X) is the same for all classes, only P $(X|C_i)$P (C_i) need be maximized. If the class a priori probabilities, P (C_i), are not known, then it is commonly assumed that the classes are equally likely, that is, P (C_1) = P (C_2) = . . . = P (C_k), and we would therefore maximize P $(X|C_i)$. Otherwise we maximize P $(X|C_i)$P (C_i). Note that the class a priori probabilities may be estimated by P (C_i) = freq(C_i, T)/|T |.

4. Given data sets with a lot of attribute, it would be computationally expensive to compute P $(X|C_i)$. In order to reduce computation in evaluating P $(X|C_i)$ P (C_i), the naive assumption of class conditional independence is made. This presumes that the standards of the attributes are provisionally independent of one an-other, given the class label of the sample. Mathematically this means that

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 6, June 2017

$$P(\mathbf{X}|C_i) \approx \prod_{k=1}^{n} P(x_k|C_i).$$

The probabilities P $(x_1|C_i)$, P $(x_2|C_i)$. . . P $(x_n|C_i)$ can easily be estimated from the training set. Recall that here x_k refers to the value of attribute A_k for sample X.

If A_k is categorical, then P $(x_k|C_i)$ is the number of samples of class C_i in T having the value x_k for attribute A_k, divided by freq (C_i, T), the number of sample of class.

(a) C_i in T .
(b) If A_k is continuous-valued, then we classically assume that the values have a Gaussian distribution with a mean μ and standard deviation σ defined by

$$g(x, \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp -\frac{(x-\mu)^2}{2\sigma^2},$$

so that

$$p(x_k|C_i) = g(x_k, \mu_{C_i}, \sigma_{C_i}).$$

We need to compute μ_{Ci} and σ_{Ci} , which are the mean and standard deviation of values of attribute A_k for training samples of class C_i.

5. In sort to calculate the class label of X, P $(X|C_i)$P (C_i) is eval-uated for each class C_i. The classifier predicts that the class label of X is C_i if and only if it is the class that maximizes P $(X|C_i)$

3.4.2 C4.5 Classifier

The resulting decision tree is generated after classification.
The classifier is trained and tested first. Then the resulting decision tree or rule set is used to classify unseen data. C4.5 is the newer version of ID3. C4.5 algorithm has many features like:

- Speed - C4.5 is significantly faster than ID3 (it is faster in several orders of magnitude)
- Memory - C4.5 is more memory efficient than ID3
- Size of decision Trees – C4.5 gets smaller decision trees.
- Rule set - C4.5 can give rule set as an output for complex decision tree.

IV. RESULTS

4.1 Implementation details

To prevent data sparsity and noise, we then derive higher-level features by condensing the data into 5 features. Such abstraction is mainly based on common knowledge of EHR data hierarchies.
In our framework, we apply several widely-used machine learning models, including Naïve Bayes & C4.5 Classifier. We perform training and evaluation on different abstraction levels of feature sets, e.g., on the 15 attributes and 5 features respectively. We conduct extensive comparison of both classifiers on the same level of features, as well as performance across the 4 different levels of phenotypes.

The Table 1 shows that the analysis of Type 1 Diabetes 500 Patients Record. When we Train patients record out of 500 patients the results shows that C 4.5 and Back propagation algorithm gives more accurate result than Naïve Bayes

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering
(An ISO 3297: 2007 Certified Organization)
Website: www.ijircce.com
Vol. 5, Issue 6, June 2017

Patients Record	Naïve Bayes	C4.5 & Back Propagation
100	72	91
200	177	195
300	277	295
400	377	395
500	477	495

Table 1 :Type 1 Diabetes Classification

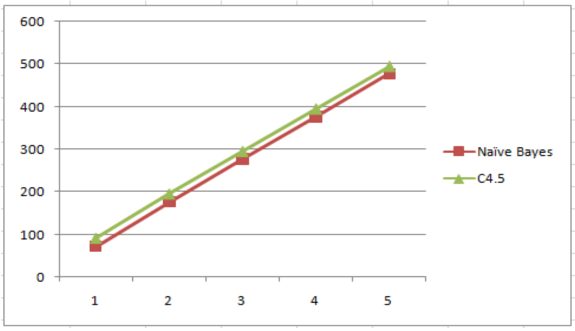

Fig 2: Line Graph for Type 1 Diabetes.

The Fig 2 shows that the result analysis of Type 1diabetes risks factor detection.

The Table 2 shows that the analysis of Type 2 Diabetes 500 Patients Record. When we Train patient's record out of 500 patients the results shows that C 4.5 and Back propagation algorithm gives more accurate result than Naïve Bayes

Patients Record	Naïve Bayes	C4.5 & Back Propagation
100	82	99
200	192	199
300	282	299
400	382	399
500	482	499

Table 2: Type 2 Diabetes classification

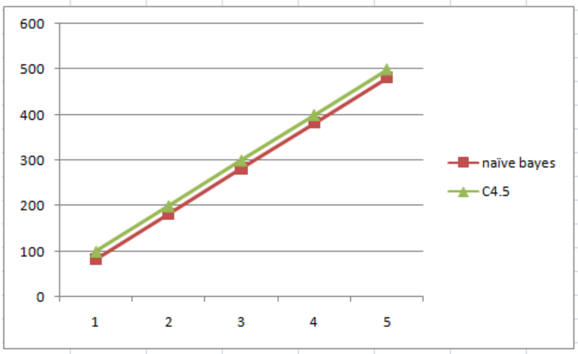

Fig 3: Line Graph for Type 2 Diabetes

The Fig 3 shows that the result analysis of Type 2 diabetes risks factor detection.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 6, June 2017

V. Conclusion

Therefore we have calculated that as the hypertriglyceridemic waist (HW) is strongly associated with type 2 diabetes Phenotype still, to date, no study has assessed the analytical power of phenotypes based on individual anthropometric measurements and triglyceride levels. The aims of the present study will be to assess the association between the HW phenotype and type 2 diabetes in Korean adults and to calculate approximately the predictive power of different phenotypes consisting of combinations of individual anthropometric measurements and triglyceride levels (TG). Between November 2006 and August 2013, 11938 subjects participated in this demonstration cross sectional study. We measured fasting plasma glucose and triglyceride levels (TG).and performed anthropometric measurements. We are working binary logistic regression (LR) to examine statistically significant differences between normal subjects and with type 2 diabetes using HW and individual anthropometric measurements. For more consistent prediction results, Naïve Bayes, two machine learning algorithms, and LR, were used to evaluate the predictive power of various phenotypes. For more reliable prediction results, two machine learning algorithms, naive Bayes (NB) and LR, were used to evaluate the predictive power of various phenotypes.

References

[1] P. T. Katzmarzyk, C. L. Crig, and L. Gauvin, “Adiposity, physical fitness and incident diabetes: The physical activity longitudinal study,” *Diabetologia*, vol. 50, no. 3, pp. 538–544, Mar. 2007.

[2] Z. Xu, X. Qi, A. K. Dahl, and W. Xu, “Waist-to-height ratio is the best indicator for undiagnosed type 2 diabetes,” *Diabetic Med.*, vol. 30, no. 6, pp. e201–e207, Jun. 2013.

3]R. N. Feng, C. Zhao, C. Wang, Y. C. Niu, K. Li, F. C. Guo, S. T. Li, C. H. Sun, and Y. Li, “BMI is strongly associated with hypertension, and waist circumference is strongly associated with type 2 diabetes and dyslipidemia, in northern Chinese adults,” *J. Epidemiol.*, vol. 22, no. 4, pp. 317–323, May 2012.

[4] A. Berber, R. G´omez-Santos, G. Fangh¨anel, and L. S´anchez-Reyes, “Anthropometric indexes in the prediction of type 2 diabetes mellitus, hypertension and dyslipidaemia in a Mexican population,” *Int. J. Obes. Relat Metab. Disorders*, vol. 25, no. 12, pp. 1794–1799, Dec. 2001.

[5] B. Balkau, D. Sapinho, A. Petrella, L. Mhamdi, M. Cailleau, D. Arondel, and M. A. Charles, D. E. S. I. R. Study Group, “Prescreening tools for diabetes and obesity-associated dyslipidaemia: Comparing BMI, waist and waist hip ratio. The D.E.S.I.R. Study,” *Eur. J. Clin. Nutr.*, vol. 60, no. 3, pp. 295–304, Mar. 2006.

[6] I. S. Okosun, K. M. Chandra, S. Choi, J. Christman, G. E. Dever, and T. E. Prewitt, “Hypertension and type 2 diabetes comorbidity in adults in the United States: risk of overall and regional adiposity,” *Obes. Res.*, vol. 9, no. 1, pp. 1–9, Jan. 2001.

[7] L. A. Sargeant, F. I. Bennett, T. E. Forrester, R. S. Cooper, and R. J.Wilks, “Predicting incident diabetes in Jamaica: the role of anthropometry,” *Obes. Res.*, vol. 10, no. 8, pp. 792–798, Aug. 2002.

[8] N. T. Duc Son le, T. T. Hanh, K. Kusama, D. Kunii, T. Sakai, N. T. Hung, and S. Yamamoto, “Anthropometric characteristics, dietary patterns and risk of type 2 diabetes mellitus in Vietnam,” *J. Amer. Coll. Nutr.*, vol. 24, no. 4, pp. 229–234, Aug. 2005.

[9] G. T. Ko, J. C. Chan, C. S. Cockram, and J.Woo, “Prediction of hypertension, diabetes,