Estimate of the HOMA-IR Cut-off Value Identifying Subjects at Risk of Insulin Resistance using a Machine Learning Approach

*Abdelhamid Abdessalem,¹ Hamza Zidoum,¹ Fahd Zadjali,² Rachid Hedjam,¹ Aliya Al-Ansari,³ Riad Bayoumi,⁴ Said Al-Yahyaee,⁵ Mohammed Hassan,⁶ Sulayma Albarwani⁷

Departments of ¹Computer Science, ²Biochemistry, ³Biology, ⁵Genetics and ⁷Physiology, Sultan Qaboos University, Muscat, Oman; ⁴Mohammed Bin Rashid University for Medicine and Health Science; Dubai, United Arab Emirates; ⁶Department of Sleep Medicine and Physiological Measurements, Canadian Health Center, Muscat, Oman

*Corresponding Author’s e-mail: ahamid@squ.edu.om

Abstract

Objective: This paper describes an unsupervised Machine Learning approach to estimate the HOMA-IR cut-off identifying subjects at risk of insulin resistance in a given ethnic group, based on the clinical data of a representative sample. Methods: We apply the approach to clinical data of individuals of Arab ancestors obtained from a family study conducted in the city of Nizwa between January 2000 and December 2004. First, we identify HOMA-IR-correlated variables to which we apply our own clustering algorithm. Two clusters having the smallest overlap in their HOMA-IR values are returned. These clusters represent samples of two populations: insulin sensitive subjects and individuals at risk of insulin resistance. The cut-off value is estimated from intersections of the Gaussian functions modelling the HOMA-IR distributions of these populations. Results: We identified a HOMA-IR cut-off value of 1.62±0.06. We demonstrated the validity of this cut-off by 1) Showing that clinical characteristics of the identified groups
match well published research findings about insulin resistance. 2) Showing a strong relationship between the segmentations resulting from the proposed cut-off and that resulting from the 2-hours glucose cut-off recommended by WHO for detecting prediabetes. Finally, we showed that the method is also able to identify cut-off values for similar problems (e.g. fasting sugar cut-off for prediabetes). Conclusion: The proposed method defines a HOMA-IR cut-off value for detecting individuals at risk of insulin resistance. Such method can identify high risk individuals at early stage which may prevent or at least delay the onset of chronic diseases like type 2 diabetes.

**Keywords:** Machine Learning; Feature Selection; K-mean++ Clustering; Insulin Resistance; HOMA-IR; T2DM.

**Advances in Knowledge**
- A machine learning method to estimate the HOMA-IR cut-off value identifying individuals at risk of insulin resistance in a given ethnic group is described.
- Identification of subjects at risk is fast and inexpensive since it consists of comparing individual’s HOMA-IR value to the defined cut-off.
- A HOMA-IR cut-off value of 1.62 is proposed to identify individuals at risk of insulin resistance among the Arab population living in Nizwa.

**Application to Patient Care**
- We recommend that identified at risk individuals undergo an additional investigation using the euglycemic clamp test for confirming or rejecting the diagnosis.
- This is of high importance for public health as it identifies high-risk individuals at an early stage which may prevent or at least delay the onset of chronic diseases such as type 2 diabetes and cardiovascular disease.

**Introduction**
Insulin resistance (IR) is a synonym for impaired insulin action,\(^1\) such as inhibition of hepatic glucose production and insulin-mediated glucose disposal.\(^2\) IR increases the incidence of metabolic syndrome (MetS), which has emerged as a major pathophysiological factor in the
development and progression of many common non-communicable diseases, including T2DM, polycystic ovary disease, dyslipidemia, hypertension, cardiovascular disease, obesity, and cancer.\textsuperscript{2, 3} Detecting this condition is therefore of high importance for public health. Several studies have illustrated a high prevalence of diabetes, impaired glucose tolerance, obesity, and hypertension among Arab populations of the Middle East.\textsuperscript{4} In Oman, the first cause of premature death is cardiovascular disease (CVD),\textsuperscript{5} and the fourth cause is type 2 diabetes mellitus (T2DM).\textsuperscript{6} Therefore, there is a need to prevent and control CVD and T2DM in Oman.

Insulin resistance is an early marker for developing these diseases; primary prevention requires the identification of high-risk individuals at an early stage. The gold standard for investigating and quantifying insulin resistance is the hyperinsulinemic (euglycemic) clamp.\textsuperscript{27} Given the complicated nature of the “clamp” technique and potential dangers of hypoglycemia in some patients,\textsuperscript{7} several surrogate estimates for insulin resistance have been proposed, such as the Homeostatic Model (HOMA-IR), Quicki and Matsuda.\textsuperscript{27, 29} HOMA-IR is the most popular among all of these indexes because of the simplicity of the underlying mathematical model (HOMA-IR= fasting glucose (mM) x fasting insulin (mU/L) /22.5). HOMA-IR has been validated to be highly correlated with the hyperinsulinemic (euglycemic) clamp ($r = 0.82$, p-value $< 0.0001$),\textsuperscript{9} and with the minimal model approximation of the metabolism of glucose ($r = -0.66$, p-value $< 0.001$).\textsuperscript{10} Several studies showed that the cut-off value for HOMA-IR depends on the ethnicity of the subjects.\textsuperscript{11, 25} Therefore, this value has to be adapted for each ethnic group to reflect the prevailing normal glycaemia.

Several statistical methods for estimating the HOMA-IR cut-off value have been reported in the literature. Most of these methods determine the cut-off values from reference ranges in which percentiles or receiver operating characteristic curves (ROC) are applied.\textsuperscript{2} The use of the percentile method analysis within a general population lacks the classification of subjects and therefore lacks sensitivity and specificity analysis. Reported ROC analyses are limited as they are based on healthy subjects and patients with T2DM or metabolic syndrome in which insulin resistance is well established in an advanced disease state. Machine Learning (ML) is a discipline of Artificial Intelligence that has proven to be very efficient in solving classification and prediction problems. The majority of ML algorithms can be categorized into two groups.\textsuperscript{28}
(1) Supervised ML in which a model is trained on a range of inputs (variables or features) that are associated with known outcomes (labels). In medicine, this might represent training a model to relate a person’s characteristics (e.g., height, weight, smoking status) to a certain outcome (onset of diabetes within five years, for example). Once the algorithm is successfully trained, it will be capable of making predictions when applied to new data. When the prediction is discrete (e.g., benign or malignant) the model is referred to as a classification while when it is of a continuous value (e.g. an individual’s life expectancy) the model is referred to as a regression. The most common supervised learning algorithms used in medicine are Decision trees, Logistic Regression, Support Vector Machine, and Artificial Neural Networks.

(2) Unsupervised learning does not involve predefined outcomes (labels) and does not need a training phase. This learning type is used to find hidden structures (or clusters) that occur within datasets. The most common methods used in this category are dimension reduction algorithms such as Principal Component Analysis (PCA), association rules generation such as ARM, and clustering algorithms such as k-means and Self-Organizing Map (SOM).

Recently, several ML-based methods for defining HOMA-IR cut-off values have been proposed. Altuve M. et al.²⁷ employed k-means clustering algorithm to perform an unsupervised classification of subjects based on unidimensional observations (HOMA-IR and the Matsuda indexes separately) and multidimensional observations (insulin and glucose samples obtained from the oral glucose tolerance test). Their results indicated that when using the HOMA-IR index alone the clusters are related to insulin resistance but when using HOMA-IR with other variables, insulin resistant cluster contains also normal samples. They argued that this could indicate that the normal samples are either developing insulin resistance or are already having the metabolic disorder. One of the drawbacks of this work is that it labels subjects with HOMA-IR greater than 2.5 as insulin resistant, which is not necessary true. Stern SE et al.³¹ developed three decision trees (decision rules) based on clinical and laboratory measurements of 2321 individuals from 17 European sites, San Antonio, Texas, and the Pima Indian reservation. The first model was based on both clinical and laboratory measurements; it achieved an AUC of 90%. The second model was developed from clinical variables only and achieved an AUC of 85%. The last one, was developed from clinical and lipid measurements and achieved an AUC of 85%. These
results indicate that insulin resistance can be defined using simple decision trees. Qu H-Q et al.\textsuperscript{11} proposed a machine learning approach to define the HOMA-IR cut-off for Americans of Mexican descent. They first identified HOMA-IR-correlated variables using two classification methods (Support Vector Machine and Logistic Bayesian Regression), then they run a clustering algorithm (k-means, with k=2) using the identified variables to obtain two reference groups representing insulin resistant and normal individuals. Finally, the sensitivity and specificity of a series of HOMA-IR cut-off values were tested against the reference groups using Matthews’s correlation coefficient (MCC). A cut-off value of 3.80 corresponding to the highest MCC value was chosen.

In this work, we propose an unsupervised machine learning method that, given a statistically representative sample of an ethnic group, it automatically estimates a HOMA-IR cut-off value defining individuals at risk of insulin resistance. This approach is applied to an Omani population living in Nizwa, but we believe that it can be used to estimate the appropriate HOMA-IR cut-off value of any other population based on clinical and biochemical variables of a statistically representative sample.

**Methods**
Data were obtained from the Oman Family study\textsuperscript{12}. The study was conducted during the period 2000-2004 and was approved by the Medical Research Committee at Sultan Qaboos University. At that time, a written and signed or thumb-printed rubber-stamped informed consent was obtained from each participant or a parent and/or legal guardian if participants were under the age of 18 years.

The dataset was obtained from a previous study on 1344 individuals of Arab ancestry as part of the Oman Family Study conducted in the Nizwa region of the Sultanate of Oman.\textsuperscript{12, 13, 35} The data includes measures of 26 variables representing anthropometric data, 24-hour systolic and diastolic blood pressures, fasting and 2-hour glucose and insulin, fasting lipid profile, hormones profile, and liver function test. Study parameters were measured using a standard clinical biochemistry lab; Plasma insulin, growth hormone , free thyroxine (FT4) and tri-iodothyronine (FT3) were measured by automated
Beckman Access 2 immuno assay system, US. Plasma leptin levels were measured using coated-tube immuno-radiometric kit (Diagnostic Systems Laboratories, USA). For routine biochemical parameters, automated Beckman Synchron CX7 (Beckman Coulter, Fullerton, CA, USA) was used to measure plasma glucose, HbA1c, total and HDL cholesterol, triglycerides and alanine transaminase (ALT). VLDL and LDL cholesterol where calculated as following: VLDL (mmol/L) = triglycerides (mmol/L) / 2.2 and LDL (mmol/L) = total cholesterol − (VLDL + HDL). Twenty-four hour ambulatory blood pressure (BP) was also measured (Schiller AG, Baar, Switzerland). Percentage total body fat was measured by electro-impedance model. Further details on data collection are described by Bayoumi RA et al, Zadjali F et al.

A data cleaning operation was conducted to remove samples with missing data or having outlier values. Details of this operation are given in the result section. The final cleaned dataset includes measurements of 26 clinical variables collected from 798 individuals (338 males and 460 females).

To determine the HOMA-IR cut-off, we adopt a three-step approach, similar to the one adopted by Qu H-Q et al. on the Americans of Mexican descent dataset, but using different techniques. First, we identify HOMA-IR correlated variables. Then we run our own clustering algorithm, to identify two reference groups representing samples of two populations: individuals at risk of insulin resistance and those with normal insulin sensitivity, respectively. Based on the assumption that HOMA-IR values of the two populations are normally distributed, we estimate the parameters of their Gaussian functions and use their intersections as an estimate of the sought cut-off value. Detailed description of this process is given in subsequent sections. There are three major differences with Qu H-Q et al. approach. 1) Their feature selection method is based on two classification techniques (Support Vectors Machine and Bayesian Logistic Regression) trained on a labelled dataset (where patients with HOMA-IR > 2.6 are labelled as insulin resistant and the rest as normal which is questionable since the objective is precisely to define this threshold), while in our method, the selection does not depend on any initial assumption about the HOMA-IR cut-off, instead, we use a statistical technique (Pearson correlation) to pre-select a subset of HOM-IR correlated variables. 2) They use a k-means algorithm to determine the two reference groups, while we have developed our own clustering algorithm (Minimum
Overlap $k$-means Clustering Algorithm) to determine the two reference groups. 3) Finally, they test several cut-off values in the range [minimum HOMA-IR value, maximum HOMA-IR value] and select the value that best fits the reference groups. Their best fit is identified using Matthews Correlation Coefficient (MCC) while our cut-off is defined by the intersection of Gaussian functions that model the HOMA-IR distributions of the two populations: individuals at risk of insulin resistance and those of normal insulin sensitivity.

In the following subsections, we provide additional details about the proposed method, indicate the assumptions we have made and our justifications for making those assumptions.

As mentioned above, the first step in our method consists of determining two reference groups, representing samples of two populations: individuals at risk of insulin resistance and those with normal sensitivity. Because insulin resistance is a complex metabolic disorder whose effect may impact the levels of several correlated clinical variables, identifying and including those variables in the analysis process might lead to an early detection of the disorder. For this reason, our clustering operation includes besides HOMA-IR, a selected subset of variables that correlate with HOMA-IR. Moreover, it’s known that feature selection is used to remove irrelevant, redundant, and noisy features. This process has several advantages. It speeds up the learning process and allows building a simpler and more accurate model, which leads to better predictors. Our feature selection algorithm consists of ranking the variables according to their correlations with HOMA-IR and eliminating those with low correlation values. The relevance of a feature $f$ to HOMA-IR is evaluated using Pearson Correlation Coefficient defined in (1):

$$
\text{corr}(f, \text{HOMA-IR}) = \frac{\text{cov}(f, \text{HOMA-IR})}{\sigma_f \sigma_{\text{HOMA-IR}}}
$$

(1)

Where $\text{cov}(x,y)$ is the covariance between features $x$ and $y$, and $\sigma_x$ is the standard deviation of feature $x$.

With the remaining features, we generate all possible combinations (subsets) and select the one that produces the best binary clustering using the $k$-means++ algorithm. The best clustering is defined as the one that minimizes the overlap between the HOMA-IR values of the resulting
clusters. We call this process, the Minimum Overlap k-means Clustering Algorithm. Since k-means is known to minimize within-cluster variances, adding the minimum overlap between the HOMA-IR means constraint, will result in a clustering having a double property: samples from the same cluster have similar clinical characteristics (k-means property) and those in two different clusters are of significantly different HOMA-IR values (minimum overlap property). Which motivate us to speculate that the two resulting clusters contain subjects with normal insulin sensitivity and those at risk of insulin resistance respectively.

The two identified clusters are just samples of the two populations: individuals at risk of insulin resistance (IR) and those of normal insulin sensitivity (NIR). We use, their respective HOMA-IR means (mean_IR, mean_NIR) and standard deviations (std_IR, std_NIR) to estimate means and standard deviations of the populations they represent as well as the 95% confidence intervals [mean_IR1, mean_IR2] and [mean_NIR1, mean_NIR2] for the estimated means. The x-coordinate (cut-off1) of the intersection point between the Gaussian functions modelled by [mean_IR1, std_IR], [mean_NIR1, std_NIR] represents the lower value of the sought cut-off point and the x-coordinate (cut-off2) of the intersection point between the Gaussian functions modelled by [mean_IR2, std_IR], [mean_NIR2, std_NIR] represents the upper value of the sought cut-off point. We set the HOMA-IR cut-off value to (cut-off1+cut-off2)/2 and the confidence interval to [cut-off1, cut-off2]. The main assumption made in these calculations, is the normality of the two populations based on the normality of large size samples from these populations.

The major processes of the proposed method consist of the following: First, a list L most-correlated variables with HOMA-IR are selected using Pearson correlation, then k-means++ algorithm (k set to 2) is applied to each subset of the list L. The subset producing clusters with the least overlap in terms of HOMA-IR values is selected and the two resulting clusters represent samples of two populations: individuals at risk of insulin resistance and individuals with normal insulin sensitivity. Finally, intersections of the Gaussian functions modelling the two populations’ HOMA-IR distributions are used to define the sought cut-off point and its confidence interval as explained earlier.
Results
All calculations are performed using Python language and related libraries including the machine learning library sklearn (https://scikit-learn.org), and the statistics library scipy.stats (https://docs.scipy.org/doc/scipy/reference/stats.html).

As mentioned earlier, the initial dataset consists of 1344 samples. First, we removed samples having missing data in any of the 26 variables. This resulted in a new dataset of 1003 samples. Then we have identified and removed the outliers (those samples with HOMA-IR values’ z-scores > 3), which resulted in a final dataset size of 798 samples having valid values for 26 variables (338 males and 460 females).

Correlation values between HOMA-IR and the 26 variables are calculated (See Table.1). Variables with correlation values less than 0.1 are removed which leads to a subset of 16 variables: fasting plasma insulin (INSU0), fasting plasma glucose (SUG0), 2-hour postprandial glucose concentration (SUG2), leptin(LEPT), waist circumference (WST), body-mass index (BMI), very-low-density cholesterol (VLDL), triglycerides (TG), percentage of fat (PERF), 24-hour diastolic and systolic blood pressures (DIAST and SYST respectively), glycated haemoglobin (HbA1c), total cholesterol (CHOL), alanine aminotransferase (ALT), human growth hormone (HGH) and free thyroxine (FT4). We have also removed SUG0 and INSU0 from the list to avoid having a clustering result dominated by these two variables as they are redundant with HOMA-IR variable that is included in the clustering process (HOMA-IR=(SUG0*INSU0)/22.5). The main benefit of the preselection step is reducing significantly the search space in the feature selection stage (from $2^{26}$ subsets to $2^{14}$ subsets) as well as reducing the noise effect that would non-correlated (and weakly-correlated) variables with HOMA-IR have on the clustering results.

We generated all possible subsets of the 14 variables (i.e.16383 subsets) to which we have applied the proposed Minimum Overlap k-means Clustering Algorithm. The variable values were codified using reference ranges consistent with WHO recommendations,\textsuperscript{16} and those of the Omani Ministry of Health.\textsuperscript{17} Figure.1 shows the HOMA-IR histograms of the two reference groups, resulting from the clustering. The first group consists of 621 individuals representing a
sample of the non-insulin resistant population; it has a HOMA-IR mean of 0.88 and a standard deviation of 0.43. The second group consists of 177 individuals representing a sample of individuals at risk of insulin resistance; it has a HOMA-IR mean of 2.71 and a standard deviation of 0.88. The optimal subset of variables producing these groups consists of HOMA-IR, VLDL, SYST, and ALT.

We have conducted normality tests on the HOMA-IR values of the two groups using the Python function (scipy.stats.normaltest()) that implements D’Agostino and Pearson’s algorithm. This algorithm combines skew and kurtosis to produce an omnibus test of normality. Obtained results (p-value = 6.74e-14 for the normal group and p-value = 2.06e-12 for the IR group), confirmed clearly the normality of the two samples’ HOMA-IR values. We used, their respective HOMA-IR means and standard deviations (0.88+/-0.43, 2.71+/-0.88) to estimate the 95% confidence intervals of the NIR and IR population means ([0.84, 0.91] and [2.58, 2.84] respectively). The x-coordinate of the intersection of the Gaussian functions modelled by [0.84, 0.43], [2.58, 0.88] is equal to 1.56. It represents the lower value of the sought HOMA-IR cut-off, and the x-coordinate of the intersection of the Gaussian functions modelled by [0.91, 0.43], [2.84, 0.88] is equal to 1.68. It represents the upper value of the sought HOMA-IR cut-off. We set the HOMA-IR cut-off to (1.56+1.68)/2 = 1.62 and the confidence interval to [1.56, 1.68]. Figure 2 shows the Gaussians’ graphs that model the two populations and their intersections.

Ideally, such work should be evaluated either by testing its performance on a labelled dataset (dataset where the samples are labelled as insulin resistant or normal) or using a longitudinal study where the validity of the predictions made by the method is checked against the evolution of the health status of the subjects. Because none of these possibilities is available to us, we adopted alternative means for the evaluation.

As an alternative way to quantitatively evaluate the performance of the method, we have applied it to solve a similar problem where labelled data are available. The problem we have chosen is the definition of the fasting glucose (SUG0) cut-off for identifying prediabetes since we can use our dataset and label the samples based on the standard cut-off value of 5.6 mM. We have first identified the SUG0-correlated variables on which we applied the Minimum Overlap k-means
**Clustering Algorithm.** The SUG0 means and standard deviations of the resulting reference groups have been used to estimate means and standard deviations of the corresponding populations and evaluate the cut-off value. The identified SUG0 cut-off was 5.89 mM and the 95% confidence interval was [5.83, 5.95]. We then segmented our dataset using both cut-off points (5.6 mM and 5.89 mM) and evaluated the level of agreement between the two classifications. Two agreement measures have been used: 24 percent agreement defined as:

\[
\frac{\text{number of agreements}}{\text{number of totals scores}}
\]  

(2)

and Kappa agreement defined as:

\[
K = \frac{Pr(a) - Pr(e)}{1 - Pr(e)}
\]  

(3)

Where \(Pr(a)\) is the observed agreement and \(Pr(e)\) is the chance agreement. The percent agreement value was equal to 0.89 and the Kappa agreement value was equal to 0.72. These results demonstrate the suitability of the proposed method in defining the SUG0 cut-off for prediabetes identification.

Chi-square test for independence shows how two sets of data are independent of each other. In our case, we applied Chi-square test to check the relationship between the segmentation resulting from the identified HOMA-IR cut-off value (1.62) and the segmentation produced by standard SUG2 cut-off value (7.8 mM) for identifying pre-diabetes. Table.2 summarizes the distribution of the samples among the four categories (Non-IR who are normal, Non-IR who are pre-diabetic or diabetic, IR who are normal and IR who are pre-diabetic or diabetic). The Python function (scipy.stats.chi2_contingency ()) returned a Chi-square statistic of 21.10, and a p-value of 4.37 e-6, which indicate a strong relationship between the two segmentations.

Table.3 shows means and standard deviations of the clinical variable values for the two groups resulting from a segmentation by a HOMA-IR threshold of 1.62. The table shows that the IR group (i.e. group at risk of insulin resistance) mean values of almost all variables are higher than NIR (i.e. the normal group) values. The only variable that shows opposite relationship is FT4. In
the next section, we will discuss these results based on the findings of research studies published in the literature.

**Discussion**

We applied the proposed machine learning approach to define the HOMA-IR cut-off value identifying individuals at risk of insulin resistance in an Omani Arab population living in Nizwa. We obtained a cut-off value of 1.62 as optimum. This value is within the range of literature-reported cut-off values 1.44 to 3.87 in different ethnic groups (Table.4).

The experiment described in section 3.4.1 demonstrates the validity of the proposed approach in identifying cut-off values in problems similar to the one under investigation.

The Chi-square test result (p-value = 4.36 e-6) indicates a strong relationship between the two segmentations: IR/NIR and normoglycemia/hyperglycemia. Table.2 shows that 89.25% of the NIR individuals (523 out of 586) are characterised with normoglycemia which what is expected from NIR individuals. It also shows that there are 162 (out of 685) individuals with normoglycemia classified in the IR category. Like M. Altuve et al, we argue that these might be the subjects who are developing insulin resistance or are already having the metabolic disorder. They should undergo a more thorough medical investigation starting with euglycemic clamp which might help in identifying IR individuals at early stage and therefore prevent or at least delay the onset of a chronic disease.

Table.3 indicates that for almost all variables, the mean values in the IR group are larger than those in the NIR group. The only variable that shows an opposite relationship is FT4. A quick review of the literature indicates an agreement of these results with the outputs of several research works. For example, a longitudinal study (of 7 years) conducted by on 1734 subjects concluded that converters to T2DM had significantly higher body mass index, waist circumference, triglyceride concentration, and blood pressure than non-converters. Insulin resistance is associated with increased triglyceride levels and VLDL, and decreased high-density lipoprotein (HDL) cholesterol. Liver function tests are frequently ordered in patients with metabolic syndrome to monitor development of non-alcoholic fatty liver disease, a study...
involving 1309 non-diabetic individuals concluded that increased m-glutamyltransferase (GGT) and alanine aminotransferase (ALT) were biomarkers of both systemic and hepatic insulin resistance.\textsuperscript{21} Also sub-clinical hypothyroidism (SCH) is associated with higher insulin levels and insulin resistance which correlates positively with TSH levels and negatively with FT3 and FT4.\textsuperscript{22} Leptin, is a body fat biomarker that can play a major predictive role of insulin resistance syndrome as a study conducted on 80 individuals showed a positive relationship between leptin and insulin resistance syndrome.\textsuperscript{23}

The current study describes an unsupervised machine learning approach for assessing insulin resistance condition given clinical data of a sample representing an ethnic group taking in consideration of the confounding variables. It allows to estimate a cut-off HOMA-IR value that identifies individuals at risk of insulin resistance. This threshold can be used as a warning signal, suggesting that subjects with HOMA-IR greater than the identified threshold should undergo the hyperinsulinemic (euglycemic) clamp to confirm or reject the prognostic. It’s the responsibility of the clinician to make sure that health condition of the patient permits conducting the proposed test. To further validate this cut-off value, a longitudinally follow-up study is warranted on healthy subjects to study the prognostic power of the cut-off value in identifying subjects who develop T2DM. Moreover, it will be interesting to apply this approach to samples of the major ethnic groups living in different parts of Oman. The output of such a research work is expected to define ethnic-dependent HOMA-IR cut-off values.

Finally, we would like to indicate the performance of the proposed approach depends on three important factors. 1) At the data collection level, samples included in the dataset should be statistically representative of the ethnic group under investigation. 2) At the implementation of the code associated with the proposed approach, there is a need to tune properly two parameters: the one identifying the outliers to remove and the one that indicates correlated variables to keep. 3) The normality (or near normality) of the IR and NIR populations.

**Conclusion**
The HOMA-IR model is a good indicator of insulin sensitivity in population studies and displays ethnic variability in the cut-off values. In this study, we identified the cut-off value of 1.62 +/-
0.06 in an Arab Omani population living in Nizwa. This approach will be helpful in future population studies concerning T2DM and MetS.

Conflict of Interest
The authors declare no conflicts of interest.

Funding
No funding was received for this study.

References


Table 1: Correlation coefficients between the variables and HOMA-IR

<table>
<thead>
<tr>
<th>Parameter</th>
<th>R</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fasting plasma insulin: INS0 (mM)</td>
<td>0.975</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Fasting blood glucose: SUG0 (mM)</td>
<td>0.442</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Plasma leptin: LEPT (ng/ml)</td>
<td>0.389</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Waist circumference: WST (cm)</td>
<td>0.337</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Body mass index: BMI (kg/m²)</td>
<td>0.322</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Percentage of fat: PERF (%)</td>
<td>0.297</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Diastolic blood pressure: DIAST (mmHg)</td>
<td>0.273</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Hemoglobin A1c: HbA1c (%)</td>
<td>0.270</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>2H- blood glucose: SUG2 (mM)</td>
<td>0.250</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Systolic blood pressure: SYST (mmHg)</td>
<td>0.228</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Plasma triglycerides: TG (mM)</td>
<td>0.206</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Very-low-density lipoprotein: VLDL (mM)</td>
<td>0.205</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Alanine aminotransferase: ALT (U/L)</td>
<td>0.189</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>
Total cholesterol: CHOL (mM) 0.145 <0.0001
Plasma cortisol: CORT (nmol/L) 0.097 0.014
Alkaline phosphatase: ALP (U/L) 0.091 0.014
Age: AGE (years) 0.076 0.0001
Thyroid hormone: TSH (mU/L) 0.057 0.091
Gender: SEX 0.012 0.800
High-density lipoprotein: HDLD (mM) -0.060 0.002
Plasma albumin: ALB (g/dL) -0.060 0.023
Immunoglobulin E: IGE (U/mL) -0.066 0.043
Total plasma proteins: TP (g/L) -0.079 0.087
Total Bilirubin: TBIL (µmol/L) -0.095 0.001
Human growth hormone: HGH(ng/mL) -0.108 0.066
Free thyroxine: FT4 (µg/dl) -0.147 <0.0001

R: Correlation coefficient. \( P \) value of correlation test.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Normal, No Diabetes (2-hour Glucose)</th>
<th>Prediabetic and Diabetic (2-hour Glucose)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal IR (below the cut off value of 1.62)</td>
<td>523</td>
<td>63</td>
</tr>
<tr>
<td>IR (above the cut off value of 1.62)</td>
<td>162</td>
<td>50</td>
</tr>
</tbody>
</table>

**Table 2:** Distribution our dataset samples among the four categories defined by the HOMA-IR cut-off value of 1.62 and SUG2 cut-off value of 7.8 mM

**Table 3:** Characteristics of the identified clusters

<table>
<thead>
<tr>
<th>Variable</th>
<th>IR mean ± std</th>
<th>NIR mean ± std</th>
<th>( P ) value</th>
</tr>
</thead>
<tbody>
<tr>
<td>INSU0 (mM)</td>
<td>9.82 ± 3.16</td>
<td>3.49 ± 1.61</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>SUG0 (mM)</td>
<td>5.84 ± 0.77</td>
<td>5.32 ± 0.54</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>SUG2 (mM)</td>
<td>6.89 ± 2.32</td>
<td>6.21 ± 1.52</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>LEPTIN (ng/ml)</td>
<td>37.67 ± 25.66</td>
<td>22.56±19.32</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>BMI (kg/m²)</td>
<td>26.96 ± 4.94</td>
<td>24.11±4.59</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>WST (cm)</td>
<td>86.34 ± 14.47</td>
<td>78.20±12.53</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>VLDL (mM)</td>
<td>0.52 ± 0.28</td>
<td>0.41 ± 0.23</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>TG (mM)</td>
<td>1.15 ± 0.62</td>
<td>0.91 ± 0.51</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>PERFAT (%)</td>
<td>26.89 ± 10.41</td>
<td>21.93± 9.65</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>
Only variables with correlation with HOMA-IR $\geq 0.1$ are shown.

<table>
<thead>
<tr>
<th>Country</th>
<th>Subjects</th>
<th>Cut off</th>
<th>Statistical Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sweden</td>
<td>Healthy population</td>
<td>2.00</td>
<td>75th percentile</td>
</tr>
<tr>
<td>France</td>
<td>Healthy population</td>
<td>3.80</td>
<td>75th percentile</td>
</tr>
<tr>
<td>Brazil</td>
<td>Healthy adult subjects</td>
<td>2.77</td>
<td>90th percentile</td>
</tr>
<tr>
<td>USA</td>
<td>Adult healthy subjects (Hispanic and non-Hispanic)</td>
<td>2.73</td>
<td>66th percentile</td>
</tr>
<tr>
<td>USA</td>
<td>Cross sectional sample adults</td>
<td>3.80</td>
<td>ML (clustering)</td>
</tr>
<tr>
<td>Portugal</td>
<td>Non-Obese non-diabetic adults</td>
<td>2.33</td>
<td>90th percentile</td>
</tr>
<tr>
<td>Iran</td>
<td>Healthy adult subjects</td>
<td>3.87</td>
<td>ROC classified by MetS</td>
</tr>
<tr>
<td>Iran</td>
<td>Cross sectional sample (healthy and diabetics adult women)</td>
<td>2.63</td>
<td>95th percentile</td>
</tr>
<tr>
<td>China</td>
<td>Healthy children and adolescents</td>
<td>3.00</td>
<td>95th percentile</td>
</tr>
<tr>
<td>Japan</td>
<td>Cross sectional sample adult non-diabetics</td>
<td>1.70</td>
<td>ROC classified by MetS</td>
</tr>
<tr>
<td>Caucasus</td>
<td>Rural population, non-diabetic</td>
<td>2.29</td>
<td>75th percentile</td>
</tr>
<tr>
<td>Thailand</td>
<td>Cross sectional</td>
<td>1.55</td>
<td>90th percentile</td>
</tr>
<tr>
<td>China (Hong Kong)</td>
<td>cross-sectional sample</td>
<td>1.44</td>
<td>75th percentile</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2.03</td>
<td>90th percentile</td>
</tr>
</tbody>
</table>

**Table 4:** Some HOMA-IR cut-offs reported in the literature $^{2, 11, 26, 30}$
**Figure 1**: HOMA-IR Histograms, IR reference group (red) and non-IR reference group (green).

**Figure 2**: The Gaussian graphs that model the two populations and used to estimate the HOMA-IR cut-off value.