

RESEARCH ARTICLE

Using Blood Indexes to Predict Overweight Statuses: An Extreme Learning Machine-Based Approach

Huiling Chen^{1,3}, Bo Yang^{2,3}, Dayou Liu^{2,3}, Wenbin Liu¹, Yanlong Liu⁴, Xiuhua Zhang^{5*}, Lufeng Hu^{5*}

1 College of Physics and Electronic Information Engineering, Wenzhou University, Wenzhou, China, **2** College of Computer Science and Technology, Jilin University, Changchun, China, **3** Key Laboratory of Symbolic Computation and Knowledge Engineering of Ministry of Education, Jilin University, Changchun, China, **4** College of Pharmaceutical Sciences, Wenzhou Medical University, Wenzhou, China, **5** Pharmaceutical Department, The First Affiliated Hospital of Wenzhou Medical University, Wenzhou, China

* wzxiuhuazhang@163.com (XZ); hulufeng79@sina.com (LH)



OPEN ACCESS

Citation: Chen H, Yang B, Liu D, Liu W, Liu Y, Zhang X, et al. (2015) Using Blood Indexes to Predict Overweight Statuses: An Extreme Learning Machine-Based Approach. PLoS ONE 10(11): e0143003. doi:10.1371/journal.pone.0143003

Editor: Dongxiao Zhu, Wayne State University, UNITED STATES

Received: May 6, 2015

Accepted: October 29, 2015

Published: November 23, 2015

Copyright: © 2015 Chen et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: This research is supported by the National Natural Science Foundation of China (NSFC) (61303113, 61133011, 61373053, 61272018 and 61402337). This work is also supported by the key construction academic subject (medical innovation) of Zhejiang Province (11-CX26), Zhejiang Provincial Natural Science Foundation of China (R1110261, LQ13G010007, LQ13F020011, LY14F020035), and the young talents project of the First Affiliated Hospital of Wenzhou Medical University (qnc 043), the open project program of Key Laboratory of Symbolic

Abstract

The number of the overweight people continues to rise across the world. Studies have shown that being overweight can increase health risks, such as high blood pressure, diabetes mellitus, coronary heart disease, and certain forms of cancer. Therefore, identifying the overweight status in people is critical to prevent and decrease health risks. This study explores a new technique that uses blood and biochemical measurements to recognize the overweight condition. A new machine learning technique, an extreme learning machine, was developed to accurately detect the overweight status from a pool of 225 overweight and 251 healthy subjects. The group included 179 males and 297 females. The detection method was rigorously evaluated against the real-life dataset for accuracy, sensitivity, specificity, and AUC (area under the receiver operating characteristic (ROC) curve) criterion. Additionally, the feature selection was investigated to identify correlating factors for the overweight status. The results demonstrate that there are significant differences in blood and biochemical indexes between healthy and overweight people (p -value < 0.01). According to the feature selection, the most important correlated indexes are creatinine, hemoglobin, hematokrit, uric Acid, red blood cells, high density lipoprotein, alanine transaminase, triglyceride, and γ -glutamyl transpeptidase. These are consistent with the results of Spearman test analysis. The proposed method holds promise as a new, accurate method for identifying the overweight status in subjects.

Introduction

The overweight condition is a rapidly growing health concern in developed and developing countries. The number of overweight and obese people has continuously increased across the world. According to a 2014 WHO (World Health Organization) survey, 39% of adults were

Computation and Knowledge Engineering of Ministry of Education, Jilin University under Grant No. 93K172013K01, the Opening Project of Zhejiang Provincial Top Key Discipline of Pharmaceutical Sciences. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

overweight. In China, the number of overweight people in the 7–18 year-old age range increased from 1.8 to 11.9% between 1985 and 2005 [1]. Being overweight raises the risk for a number of diseases, such as high blood pressure, blood glucose, dyslipidemia, diabetes mellitus, coronary heart disease, certain forms of cancer, stroke, and sleep-breathing disorders [2–5]. Consequently, being able to identify the overweight status is an important public health concern.

To date, several machine learning methods have been proposed to automatically diagnose subjects' body mass index (BMI). Most have exhibited promising accuracy. Hardalac et al. [6] developed neuro-fuzzy systems that included NEFCLASS and CANFIS; they analyzed medical parameters such as subjects' BMIs and divergent arteries. The results showed that the classification performance of NEFCLASS was better than that of CANFIS, and that obesity affects BMI rather than divergent arteries. Zhang et al. [7] proposed using logistic regression and six data mining methods to identify overweight versus obese children at different ages using the parameters of children from birth to around 3 years old. The results showed that the incorporation of non-linear interactions enhanced classification performance, and data mining methods such as artificial neural networks (ANN) and Bayesian methods perform better than logistic regression. Ergun [8] proposed the ANN and logistic regression-based automatic diagnosis systems for identifying obese from non-obese subjects using artery parameters and BMI. The results demonstrated that the classification performance of ANN was better than logistic regression, and BMI is more affected than divergent arteries in obese subjects. Valavanis et al. [9] developed hybrid ANN-based methods, including PDM-ANN and GA-ANN, for identifying overweight versus healthy subjects using gender, genes, and nutrition habits as parameters. These methods showed promising generalization ability; important differentiating factors were identified, including gender, six genetic variations, and 18 nutrition-related variables. Heydari et al. [10] proposed ANN and logistic regression for classifying normal and obese classes using subjects' demographics, lifestyle information, and anthropometric measurements. The results showed that both ANN and logistic regression are effective classifiers for obesity detection, and they are not significantly different in capabilities. Recently, Lee et al. [11] predicted BMI status (normal, overweight, and obese) by logistic regression, bagging and random forest algorithms. Based on the combination of voice features, the results showed that the classification models built by logistic regression were better than the other two algorithms.

As can be seen from these studies, ANN is a popular method for detecting subjects' BMI statuses; it achieves promising classification accuracy among the available machine learning methods. However, traditional ANN training algorithms such as the back propagation method can become trapped in local minima because they are gradient descent-based algorithms. Another disadvantage of ANN is that there are many network parameters that need to be carefully specified. Unlike ANN, the extreme learning machine (ELM) [12] chooses input weights and hidden biases randomly, and output weights are analytically determined using the Moore-Penrose generalized inverse. As a result, the ELM learns more quickly and keeps fewer fine-tuning parameters than ANN, while maintaining excellent generalization performance. ELM has been applied in fields such as disease diagnosis [13, 14], image quality assessment [15], face recognition [16], land cover classification [17] and hyperspectral images classification [18]. To the best of the authors' knowledge, ELM has not yet been used in overweight modeling applications.

In addition, related works on overweight detection reveal that there is no scientific literature that explores overweight modeling using blood samples. Blood examination is standard in the medical field, used to evaluate liver function, renal function, and blood-lipid and glucose levels. This study is the first to explore the potential of ELM in overweight modeling using information from blood samples. Moreover, this study is the first attempt to identify the most

informative and influential factors that discriminate overweight from healthy subjects. For comparison purposes, ANN, the most widely used model in overweight modeling and the advanced support vector machines [19] (SVM), were also used for overweight modeling. In addition, feature selection was employed to identify correlating factors. The Fisher Score selection was employed for pre-processing before the classification models were constructed. The classification accuracy (ACC), AUC, sensitivity, and specificity of the proposed approach were examined by comparing the results against the real-life dataset. Samples were collected from The First Affiliated Hospital of Wenzhou Medical University. Promisingly, the developed ELM-based approach achieved 90.54% ACC, 90.17% AUC, 83.54% sensitivity, and 96.80% specificity.

In summary, this paper makes the following contributions: (1) It proposes a new perspective for overweight modeling using blood samples; (2) It proposes an effective and efficient approach for automatically identifying the overweight status; and (3) The most relevant indexes were incrementally detected with the aid of the feature selection.

Materials

1. Ethics Statement

The Medical Ethics Committee of The First Affiliated Hospital of Wenzhou Medical University approved of the study, which was conducted in accordance with the Declaration of Helsinki. As a retrospective investigation, no written consents were obtained from participants, and all individual participant information was collected anonymous. Samples and participant data were provided by the information department of The First Affiliated Hospital of Wenzhou Medical University. Participants with significant cardiovascular, hepatic disease, renal disease, uncontrolled thyroid disease, significant central nervous system-related, or psychiatric disorders were excluded from this study. The ethics committees waived the need for written informed consent and approved this consent procedure.

2. Data Preparation

A stadiometer was used to measure height, and a self-zeroing weight scale was used to measure weight. Participants wore foam slippers, and a paper shirt and pants during the weighing process. There were 179 male and 297 female participants in this study. According to the WHO definition, participants were included in this study if their BMI was 25.0–30 kg/m², and defined as overweight [20]. Samples were divided into healthy people ($n = 251$) and overweight people ($n = 225$); their BMIs were 21.79 ± 0.92 kg/m² and 26.17 ± 0.84 kg/m², respectively. Fasting venous blood samples of all participants were collected into separation gel tubes, and blood and biochemical indexes were determined using a Hitachi 705/717 instrument at the physical examination center of the hospital. A total of 18 blood indexes and 16 biochemical indexes were included in this study; Table 1 includes their detailed descriptions, as well as age, red blood cell count, and white blood cell in urine. All these data have been uploaded as an attachment in supplementary file which can be accessed freely.

Statistical analysis was performed using SPSS 17 software. The BMI, age, blood, and biochemical indexes of the two groups were analyzed by a One-Way ANOVA test to discover statistical differences. Table 2 lists detailed statistical descriptions. The correlation between BMI with blood and biochemical indexes were analyzed using a Spearman test. The p-values that were less than 0.05 (the 5% significance level) were considered to indicate statistical significance in all analyses.

Table 1. All of the features used in this study and their brief descriptions.

Feature	Brief description	Feature	Brief description
X ₁	Age	X ₂₀	absolute value of monocyte (AVM)
X ₂	triglyceride (TG)	X ₂₁	red blood cell (RBC)
X ₃	glucose (GLU)	X ₂₂	red blood cell in urine (RBCU)
X ₄	low density lipoprotein (LDL)	X ₂₃	hematokrit (HCT)
X ₅	high density lipoprotein (HDL)	X ₂₄	percentage of leukomonocyte (PLC)
X ₆	total cholesterol (CHO)	X ₂₅	absolute value of leukomonocyte (AVLC)
X ₇	alanine transaminase (ALT)	X ₂₆	mean corpuscular volume (MCV)
X ₈	aspartate aminotransferase (AST)	X ₂₇	mean corpuscular hemoglobin (MCH)
X ₉	γ-glutamyl transpeptidase (γ-GT)	X ₂₈	mean corpuscular hemoglobin concentration (MCHC)
X ₁₀	total protein (TP)	X ₂₉	mean platelet volume (MPL)
X ₁₁	albumin (ALB)	X ₃₀	absolute value of eosinophils (AVE)
X ₁₂	creatinine (CR)	X ₃₁	percentage of eosinophils (PE)
X ₁₃	urea nitrogen (BUN)	X ₃₂	hemoglobin (HB)
X ₁₄	alkaline phosphatase (AKP)	X ₃₃	blood platelet (PLT)
X ₁₅	total bilirubin (TBIL)	X ₃₄	thrombocytocrit (THR)
X ₁₆	direct bilirubin (DBIL)	X ₃₅	percentage of neutrophils (PN)
X ₁₇	uric acid (UA)	X ₃₆	absolute value of neutrophils (AVN)
X ₁₈	white blood cell in urine (WBCU)	X ₃₇	red cell volume distribution width (RBCVD)
X ₁₉	percentage of monocyte (PMC)		

doi:10.1371/journal.pone.0143003.t001

Methods

1. Fisher Score

The Fisher Score [21] is one of the most commonly used and efficient supervised feature weighing methods. It determines the most discriminative features according to the fisher criterion. Given the dataset of n instances $\{x_i, y_i\}$, where $x_i \in \mathfrak{R}^m$ represents that the input feature space has m features and $y_i \in \{1, 2, \dots, c\}$ is the corresponding class labels. The score of the m -th feature can be directly measured as follows:

$$F_m = \frac{\sum_{i=1}^c n_i (\mu_i^m - \mu^m)^2}{(\sigma^m)^2} \tag{1}$$

where $(\sigma^m)^2 = \sum_{i=1}^c n_i (\sigma_i^m)^2$, n_i denotes the number of instances in class i , μ_i^m and μ^m represent the mean value of class i and the global mean value corresponding the m -th feature, respectively. σ_i^m and σ^m mean the variance of class i and the global variance corresponding the m -th feature, respectively. According to Eq (1), the larger score value represents that the m -th feature has more discriminative power between the different classes.

2. Extreme Learning Machine (ELM)

This section provides a brief description of ELM; refer to [12, 22] for more information. Given a training dataset with N samples $\mathfrak{R} = \{x_i, t_i\}_{i=1}^N$, $x_i \in R^n$ is the input feature vector with n features, and $t_i \in R^m$ represents the target vector with m dimensions. The output of ELM can be written as follows [12]:

$$\sum_{i=1}^k \beta_i g(w_i \cdot x_j + b_i) = o_j, j = 1, 2, \dots, N \tag{2}$$

Table 2. Characteristics of blood and biochemical parameters in healthy and overweight people.

Index	Overweight (n = 225)				Healthy (n = 251)				p-value
	Mean	SD	Min	Max	Mean	SD	Min	Max	
AGE	44.73	10.41	22.00	76.00	42.90	10.41	24	82	0.057
TG	2.20	1.80	0.46	14.17	0.98	0.64	0.23	5.73	0.000
GLU	5.92	1.24	4.50	17.00	5.44	0.88	4.20	15.20	0.000
LDL	2.87	0.71	1.28	5.85	2.57	0.76	0.76	5.82	0.000
HDL	1.23	0.28	0.69	2.21	1.61	0.32	0.88	2.72	0.000
CHO	5.00	0.91	2.98	9.97	4.69	0.90	2.54	8.25	0.000
ALT	35.66	23.31	4.00	137.00	16.46	10.01	2.00	69.00	0.000
AST	25.13	9.84	10.00	83.00	19.37	5.91	10.00	49.00	0.000
γ-GT	56.95	65.98	10.00	564.00	15.95	10.57	6.00	121.00	0.000
TP	76.65	3.50	66.00	86.60	76.14	3.80	66.20	85.50	0.130
ALB	47.69	2.66	38.80	54.50	46.48	2.38	40.80	53.30	0.000
CR	72.46	13.41	41.00	107.00	51.92	7.01	34.00	78.00	0.000
BUN	5.28	1.22	2.80	9.10	4.73	1.17	2.30	8.50	0.000
AKP	82.71	20.65	32.00	170.00	65.35	19.91	23.00	185.00	0.000
TBIL	11.82	4.13	5.00	27.00	9.82	3.20	5.00	22.00	0.000
DBIL	3.76	1.45	2.00	11.00	3.10	1.06	2.00	7.00	0.000
UA	381.48	83.29	98.00	601.00	267.31	50.55	128.00	403.00	0.000
WBCU	11.27	25.04	0.00	203.10	31.67	56.41	0.30	474.50	0.000
PMC	0.07	0.02	0.02	0.14	0.07	0.02	0.03	0.14	0.005
AVM	0.50	0.17	0.00	1.30	0.39	0.12	0.20	0.80	0.000
RBC	5.04	0.48	3.75	6.87	4.46	0.38	3.45	6.03	0.000
RBCU	54.01	473.50	0.00	6977.40	42.96	116.47	1.20	1500.00	0.721
HCT	0.46	0.04	0.31	0.53	0.41	0.03	0.29	0.46	0.000
PLC	0.37	0.07	0.15	0.59	0.37	0.08	0.11	0.60	0.781
AVLC	2.47	0.70	0.00	5.40	2.11	0.57	0.90	4.60	0.000
MCV	91.04	6.23	62.20	108.20	91.22	6.45	56.50	104.30	0.757
MCH	30.17	2.53	18.90	34.80	29.40	2.48	15.40	33.10	0.001
MCHC	331.01	11.65	268.00	357.00	321.98	9.66	272.00	345.00	0.000
MPL	10.83	1.91	0.00	14.10	10.71	2.17	0.00	14.20	0.521
AVE	0.17	0.12	0.00	0.60	0.13	0.10	0.00	0.90	0.000
PE	0.02	0.02	0.00	0.10	0.02	0.01	0.00	0.11	0.011
HB	151.59	15.03	83.00	179.00	130.51	9.63	78.00	149.00	0.000
PLT	223.42	49.68	102.00	437.00	226.57	51.96	85.00	451.00	0.500
THR	0.24	0.06	0.00	0.45	0.24	0.07	0.00	0.42	0.772
PN	0.53	0.08	0.34	0.78	0.54	0.08	0.30	0.84	0.403
AVN	3.67	1.25	0.00	9.60	3.16	1.04	1.30	7.30	0.000
RBCVD	12.85	0.98	11.40	19.30	12.90	1.10	11.60	20.90	0.536

doi:10.1371/journal.pone.0143003.t002

where $g(x)$ is the activation function, k is the number of hidden neurons, β_i is the weight vector between the i th hidden neuron and the output layer, w_i is the weight vector between the i th neuron in the hidden layer and the input layer, and b_i indicates the bias of the i th neuron in the hidden layer, o_j is the target vector of the j th input data. If ELM can approximate these N samples with zero error, we can get $\sum_{i=1}^k \beta_i g(w_i \cdot x_j + b_i) = t_j, j = 1, 2, \dots, N$. The above equation

can be reformulated as follows:

$$H\beta = T \tag{3}$$

where H [23] represents the hidden layer output matrix of the neural network:

$$H = \begin{pmatrix} g(w_1 \cdot x_1 + b_1) & \dots & g(w_k \cdot x_1 + b_k) \\ \vdots & \ddots & \vdots \\ g(w_1 \cdot x_N + b_1) & \dots & g(w_k \cdot x_N + b_k) \end{pmatrix}_{N \times k} \tag{4}$$

$\beta = [\beta_1, \dots, \beta_k]^T$ is the matrix of output weights from the hidden layer to the output layer, and $T = [t_1, \dots, t_N]^T$ denotes target labels' vectors. Under the assumption that [24, 25] the input weights and the hidden layer biases of single hidden layer feed-forward neural network (SLFNs) can be arbitrarily given, the output weights β can be analytically determined by the Moor-Penrose (MP) generalized inverse of matrix H, as is shown in the following Equation:

$$\beta = H^\dagger T \tag{5}$$

By using the MP inverse method, the ELM generalization performance can be achieved with a dramatically increased learning speed [22].

3. The Proposed Method

This section briefly describes the proposed method. The proposed method provides an efficient and accurate prediction tool for distinguishing overweight from healthy people using blood indexes. Fig 1 illustrates the flowchart of the proposed method. First, feature selection identifies the informative blood indexes. An incremental feature selection procedure is then performed to find the most representative feature subsets with top ranked features. During this procedure, features in the ranked feature list are added one-by-one from high to low rank. A total of 37 feature sets are constructed; the set that yields the best classification performance is considered the optimal feature set. Finally, the obtained optimal feature set is fed to the ELM model for performance evaluation. Before the ELM model is evaluated, it is important to first address the selection of hidden neurons and activation functions. To determine the optimal ELM structure, both issues are handled via the 10-fold cross validation (CV) analysis. The Sigmoid function (sig), Sine function (sin), Hard-limit function (hardlim), Triangular basis function (tribas) and Radial basis function (radbas) are used for evaluation. Different hidden neurons of 5, 20, 35, 50, 65, 80 and 95 are chosen for comparison. The pseudocode of the proposed method is given below.

```

Begin
  For j = 1:k
    Training set ← k-1 subsets;
    Validation set ← remaining subset;
    Rank features incrementally using the Fisher Score
     $F_i = \{f_1, \dots, f_i, \dots, f_{37}\} (1 \leq i \leq 37)$ ;
    Train the ELM classifier on each feature subset  $f_i$  with the top  $i$  ranked
    features using a variation of the number of hidden neurons and type of
    activation functions;
    Evaluate the trained ELM model on the validation set with the correspond-
    ing reduced feature set;
  EndFor;
  Return the average classification accuracy rates of ELM over the  $j$ th val-
  idation set;
End.

```

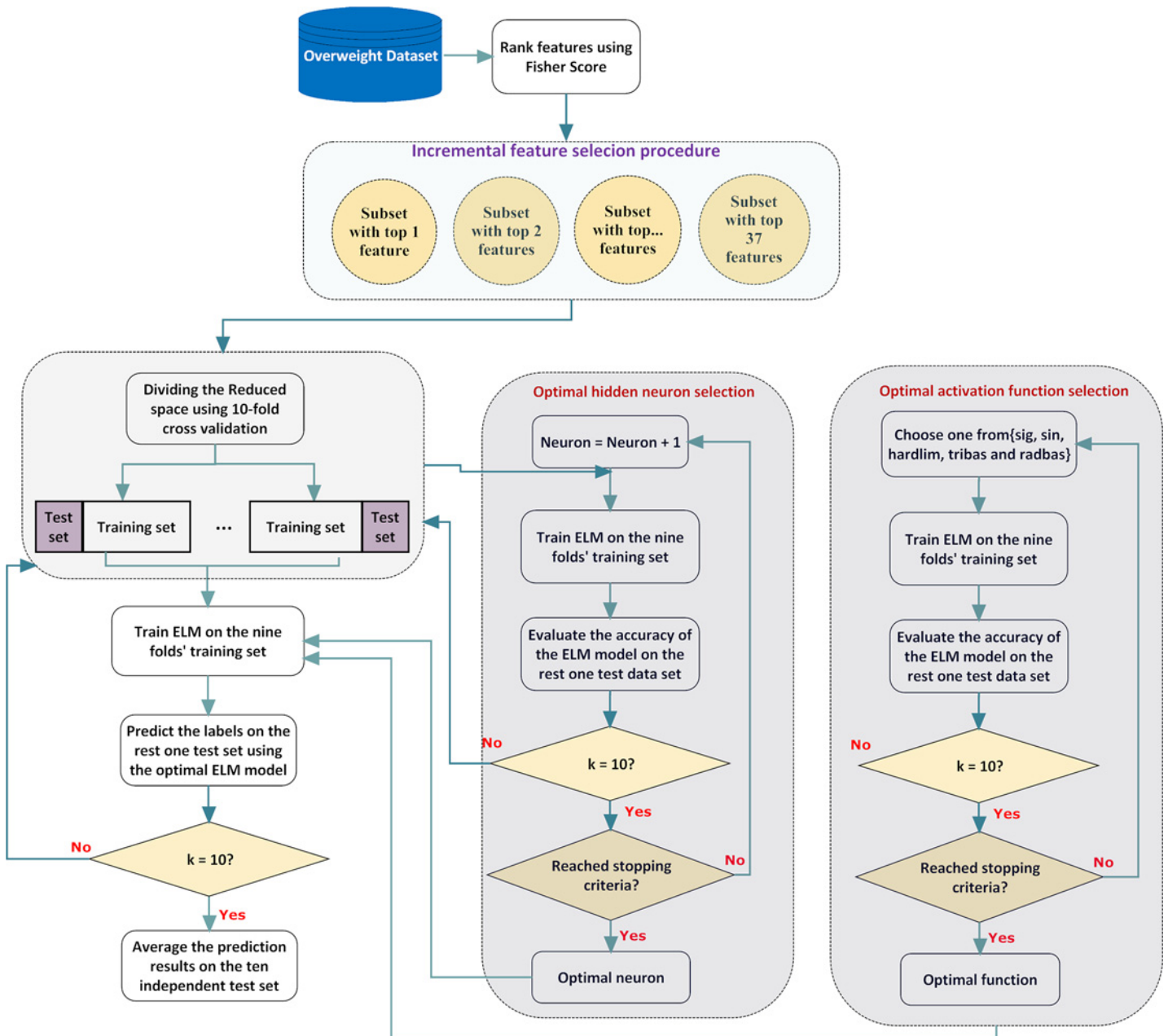


Fig 1. Overall procedure of the proposed method.

doi:10.1371/journal.pone.0143003.g001

4. Experimental Designs

4.1 Experimental Setup. To verify the proposed ELM approach, the state-of-the-art SVM and the commonly used ANN method in overweight modeling were employed for comparison. The famous back propagation neural network (BPNN) with the Levenberg-Marquardt training algorithm in the MATALAB neural network toolbox was adopted. Implementation code available at <http://www3.ntu.edu.sg/home/egbhuang> was used to construct the ELM model. For SVM, the LIBSVM toolbox developed by Chang and Lin [26] was adopted. The Fisher Score feature selection method was implemented from scratch in the MATALAB.

Data was scaled into the range $[-1, 1]$ before classification. The empirical experiment was conducted on an AMD Athlon 64 X2 Dual Core Processor 5000+ (2.6 GHz) with 4GB of RAM running Windows 7.

4.2 Data Division. The k-fold CV [27] was used to evaluate classification performance to guarantee unbiased results. The value of k is often set to 10 in the literature. As a result, whole data samples will be randomly split into 10 subsets; each time, nine subsets are used for training and the remaining one is used as the test set. The process ran 10 times. The final result was computed by averaging the result across all 10 trials. It should be noted that it is more reasonable to keep the same proportion of the samples in each fold as that of the entire dataset when splitting the data; therefore, the above stratified k-fold CV strategy is employed for analysis in the following experiment.

4.3 Evaluation Criteria. To evaluate the proposed method, commonly used evaluation criteria such as classification accuracy (ACC), the area under the receiver operating characteristic curve (AUC) [28], sensitivity, and specificity were analyzed. They are defined as follows:

$$ACC = \frac{TP + TN}{TP + FP + FN + TN} \times 100\% \quad (6)$$

$$Sensitivity = \frac{TP}{TP + FN} \times 100\% \quad (7)$$

$$Specificity = \frac{TN}{FP + TN} \times 100\% \quad (8)$$

where TP, FN, TN, and FP are the number of true positives, false negatives, true negatives and false positives, respectively. AUC is one of the most popular methods for evaluating the performance of the binary classifier. A perfect classifier provides an AUC of 1. This study adopted the AUC algorithm developed in [29].

Results

1. ELM Classification Performance

Previous studies [14, 30] showed that the activation functions and hidden neurons have more or less impact on ELM performance. Therefore, these two factors were investigated in the following experiment. The influence of different activation functions on the performance of the ELM model was investigated. Five activation functions including sig, sin, hardlim, tribas, and radbas were used. Fig 2 displays the classification accuracy of ELM with different activation functions from the function of the different number of neurons. The ELM with the sig activation function outperforms ELM with other functions. Therefore, the Sigmoid function was employed in subsequent experiment analysis.

To determine the optimal number of hidden neurons, the validation accuracy as the function of the number of hidden neurons was recorded. As shown in Fig 3, the performance of ELM is relatively stable with the increase of hidden neurons. It is therefore necessary to determine the most suitable number of hidden neurons for ELM. Therefore, different models were built with different given hidden neurons of 5, 20, 35, 50, 65, 80 and 95. Table 3 presents the average classification performance results of 10-fold CV with different numbers of hidden neurons. As can be seen from the table, the classification performance of ELM models varied with differing numbers of hidden neurons. 35 hidden neurons achieved the highest validation accuracy. Therefore, 35 hidden neurons were chosen to create the training model in the following analysis. After the activation function and the number of hidden neurons were determined, the

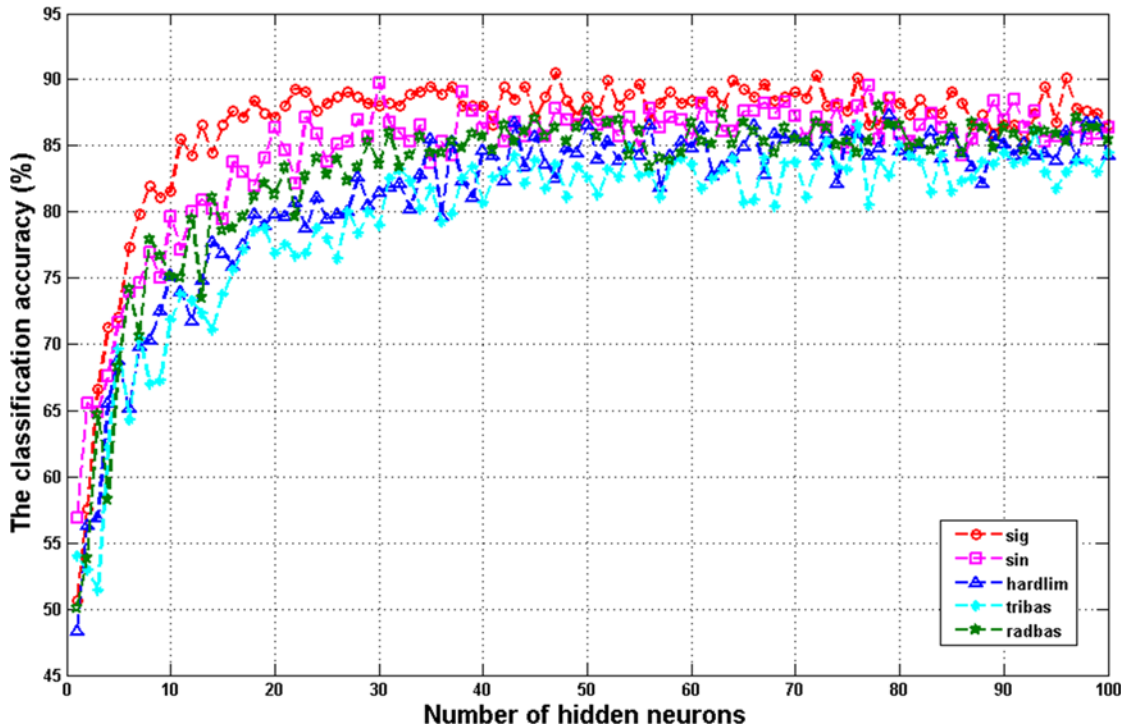


Fig 2. The relationship between ELM with different activation functions and the different number of hidden neurons.

doi:10.1371/journal.pone.0143003.g002

final model trained for prediction. The random input weights and the hidden layer biases acquired in this study are listed in Information [S1 Table](#). [Table 4](#) displays the detailed results of the 10-fold CV of the ELM. From the table, it can be seen that the ELM model achieves high performance with average results of 90.32% ACC, 89.98% AUC, 83.95% sensitivity, and 96.02% specificity.

2. Comparison with SVM and BPNN

To verify the effectiveness of the ELM model, the SVM with RBF kernel and BPNN were implemented for comparison in the whole feature space on the same dataset. For SVM, a grid-search technique [31] was employed using 10-fold CV to determine the optimal parameter values of the RBF kernel function. The range of related parameters C and γ varied between $C = \{2^{-5}, 2^{-3}, \dots, 2^{15}\}$ and $\gamma = \{2^{-15}, 2^{-13}, \dots, 2^1\}$. 99 parameter combinations of (C, γ) were tried (Training accuracy surface of SVM with parameters obtained by grid search is listed in Information [S1 Fig](#)); the one with the best CV accuracy was chosen as the parameter value of the RBF kernel. Then, the best (C, γ) parameter pair was used to create the training model. Concerning BPNN, the three layer BP network was used, and different settings of the number of nodes in the hidden layers (5, 10, 15, 20, 25 and 30) and the different learning epochs (50, 100, 200 and 300) were tried as training stopping criteria. According to the preliminary simulation results, the best result was obtained with the hidden nodes of 10 and the learning epoch of 200. These parameter settings were used for the subsequent analysis.

[Fig 4](#) compares the ACC, AUC, sensitivity, and specificity of ELM, SVM, and BPNN. ELM performed the best among the three methods in terms of ACC, AUC, and sensitivity. Compared to ELM, SVM achieved slightly higher specificity, and BPNN obtained similar sensitivity. [Table 5](#) also presents the confusion matrices obtained by the three methods via 10-fold CV

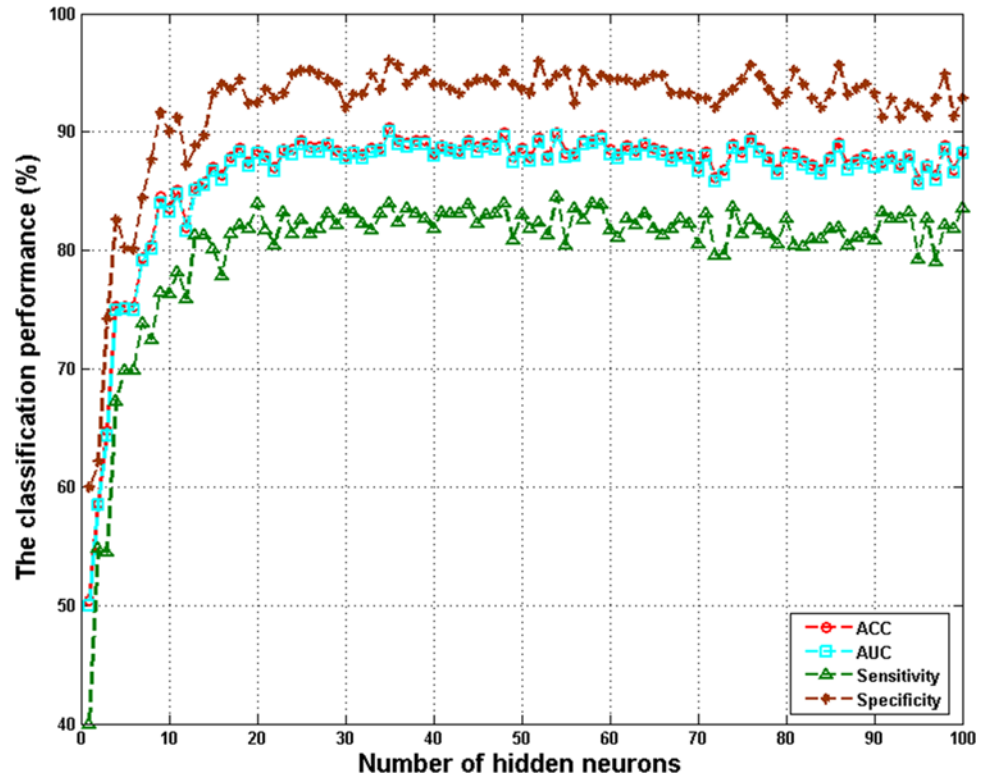


Fig 3. ACC, AUC, sensitivity, and specificity versus different hidden neuron numbers for ELM via 10-fold CV.

doi:10.1371/journal.pone.0143003.g003

analysis. As can be seen from Table 5, ELM correctly recognized 189 out of 225 overweight cases and 241 out of 251 healthy cases. It misclassified 36 overweight cases as healthy cases and 10 healthy cases as overweight cases. SVM correctly classified 184 out of 225 overweight cases and 244 out of 251 healthy cases. It misclassified 41 overweight cases and 7 healthy cases. BPNN correctly recognize 188 out of 225 overweight cases and 211 out of 251 healthy cases. It misclassified 37 overweight cases and 40 healthy cases. Based on these results, ELM and SVM outperformed BPNN, especially concerning the capability to recognize healthy cases. Additionally, ELM is more effective than SVM and BPNN at discriminating overweight from healthy

Table 3. Average classification performance results of 10-fold CV.

Different number of hidden neurons	Average classification performance			
	ACC (%)	AUC (%)	Sensitivity (%)	Specificity (%)
5	76.69	76.28	68.85	83.71
20	88.44	88.02	80.81	95.23
35	90.32	89.98	83.95	96.02
50	88.87	88.56	82.71	94.42
65	87.61	87.28	81.74	92.83
80	87.39	87.10	82.15	92.05
95	87.80	87.54	82.25	92.83

The best results have been shown in bold.

doi:10.1371/journal.pone.0143003.t003

Table 4. The detailed results obtained by the ELM model.

Fold	ACC (%)	AUC (%)	Sensitivity (%)	Specificity (%)
1#	89.36	88.91	81.82	96.00
2#	91.67	91.48	86.96	96.00
3#	87.50	87.30	82.61	92.00
4#	91.67	91.48	86.96	96.00
5#	91.67	91.30	82.61	100.00
6#	93.75	93.65	91.30	96.00
7#	91.67	91.26	86.36	96.15
8#	87.23	86.36	72.73	100.00
9#	89.36	89.18	86.36	92.00
10#	89.36	88.91	81.82	96.00
Avg.	90.32	89.98	83.95	96.02
Dev.	2.09	2.23	4.97	2.67

doi:10.1371/journal.pone.0143003.t004

controls. These results verify that ELM has better generalization capabilities than other counterparts. The CPU time needed for the training procedure of ELM, SVM, and BPNN were also recorded in the experiment. SVM required 95 seconds, BPNN required 169 seconds, and ELM required 0.5 seconds. ELM is far more efficient than SVM and BPNN for overweight recognition.

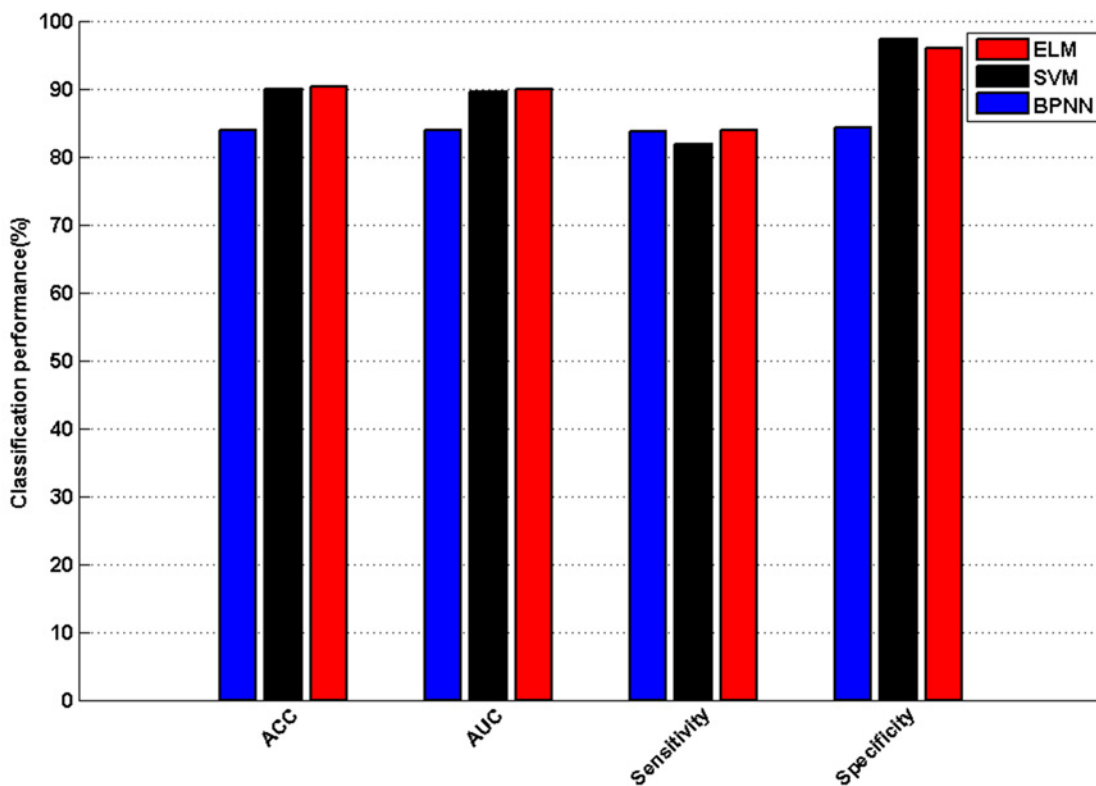


Fig 4. Comparison results of classification performance among ELM, SVM, and BPNN.

doi:10.1371/journal.pone.0143003.g004

Table 5. The confusion matrix obtained by ELM, SVM and BPNN via 10-fold CV.

ELM	Predicted overweight people	Predicted healthy controls
Overweight people	189	36
Healthy controls	10	241
SVM		
Overweight people	184	41
Healthy controls	7	244
BPNN		
Overweight people	188	37
Healthy controls	40	211

doi:10.1371/journal.pone.0143003.t005

3. Classification Results Based on Feature Selection

To identify the most relevant factors determining the overweight status, feature selection was implemented to rank features before classification. Based on the Fisher Score feature selection algorithm, a ranked feature list was obtained on the basis of each feature’s relevance to the sample class. As shown in Fig 5, the importance of each index is computed; these lists of ranked features were used to establish the optimal feature subset in the incremental feature selection procedure. 37 different feature subsets can be obtained by adding features one-by-one from higher to lower rank according to importance. The ELM classifier is accordingly built for each feature subset and incrementally evaluated by 10-fold CV. Fig 6 plots the incremental feature selection curve, which reveals the relationship between the classification accuracy and the feature subset. The classification performance obtained on the top ranked five, seven and nine features is shown in Information S2 Table. The maximal classification accuracy is obtained at 90.54% when the top nine ranked features comprise the feature set. These nine features corresponded to the nine blood indexes of CR, HB, HCT, UA, RBC, HDL, ALT, TG and γ -GT. These nine features were selected as the optimal feature set to construct the classifier. The hierarchical clustering heat map is developed based on the nine selected features as shown in Information S2 Fig. The optimal feature set results in an average of 90.54% ACC, 90.17% AUC, 83.54% sensitivity, and 96.80% specificity for ELM. Fig 7 shows the comparison results from ELM without feature selection versus ELM with the optimal feature set in terms of ACC, AUC, sensitivity, and specificity via 10-fold CV analysis. The ACC, AUC, and specificity of the optimal feature set are all superior to those of the whole feature set. The sensitivity of the optimal feature set is also comparable to that of the original feature set. Table 6 lists the results of the ELM confusion matrix on the whole feature set and the optimal feature set. ELM with the optimal feature set and ELM with the original feature set had almost the same capability to discriminate overweight cases from the healthy controls. These results suggest that the original feature set contained redundant or irrelevant information. It also indicates that the Fisher Score feature selection has the capability to identify the most informative features.

Discussion

Routine blood tests include evaluations of liver function, renal function, and blood-lipid and glucose levels. This data can identify the physiological state of a subject. Overweight subjects have different metabolic activity compared to healthy subjects, resulting in detectable differences. A clinical study on a teenage population to study obesity-related alterations in laboratory parameters confirmed that obese subjects have systematical changes in blood test parameters

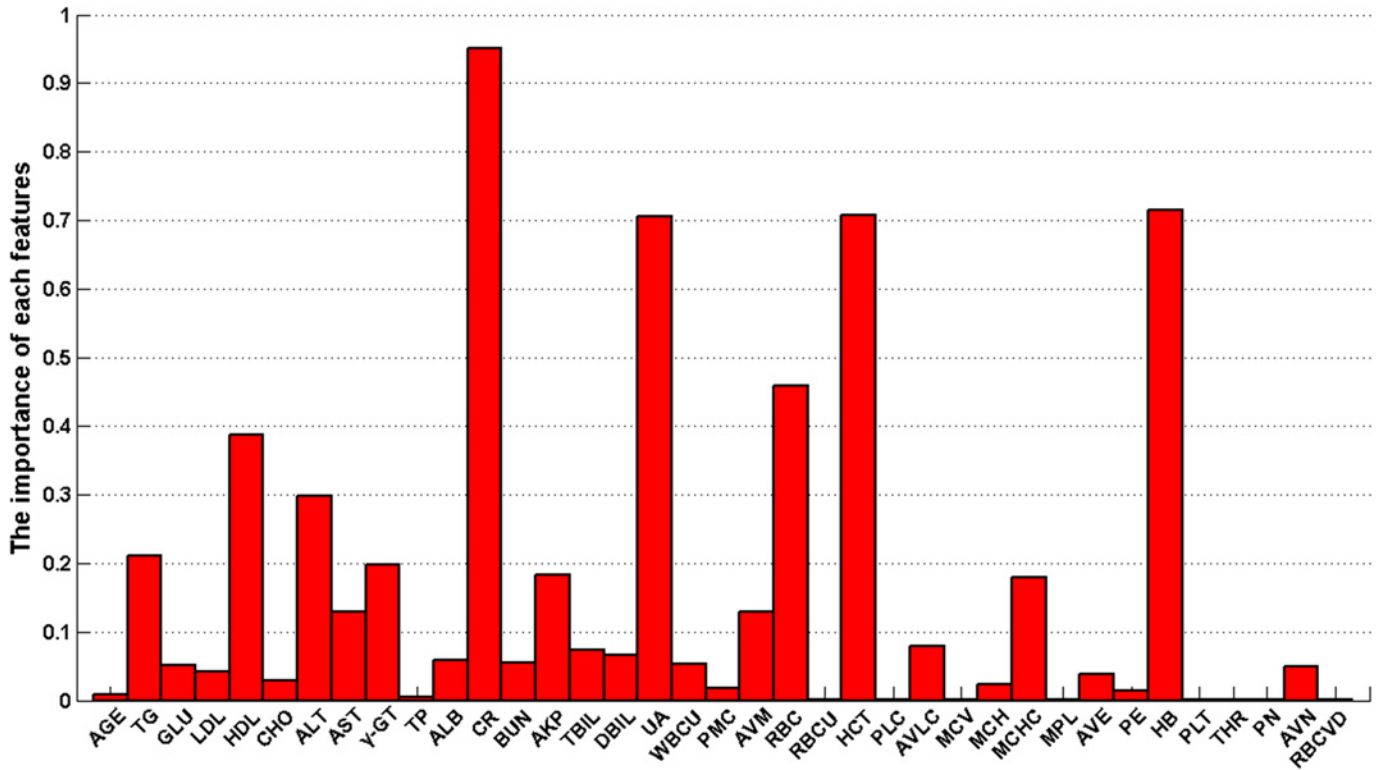


Fig 5. The importance of each index obtained by the Fisher Score feature selection.

doi:10.1371/journal.pone.0143003.g005

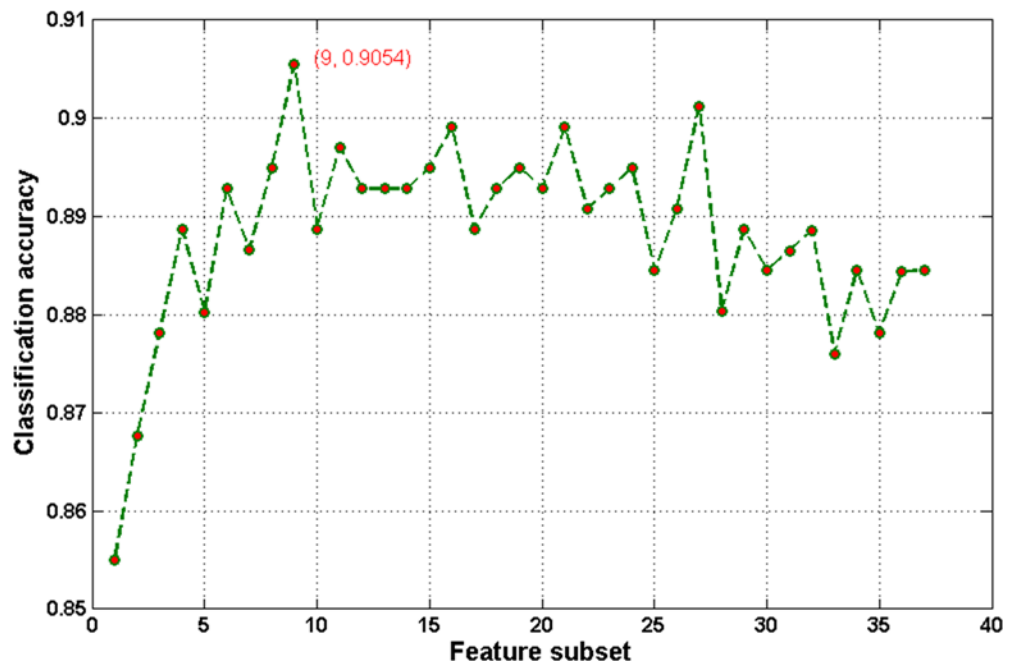


Fig 6. The incremental feature selection curve: the classification accuracy against the feature subset.

doi:10.1371/journal.pone.0143003.g006

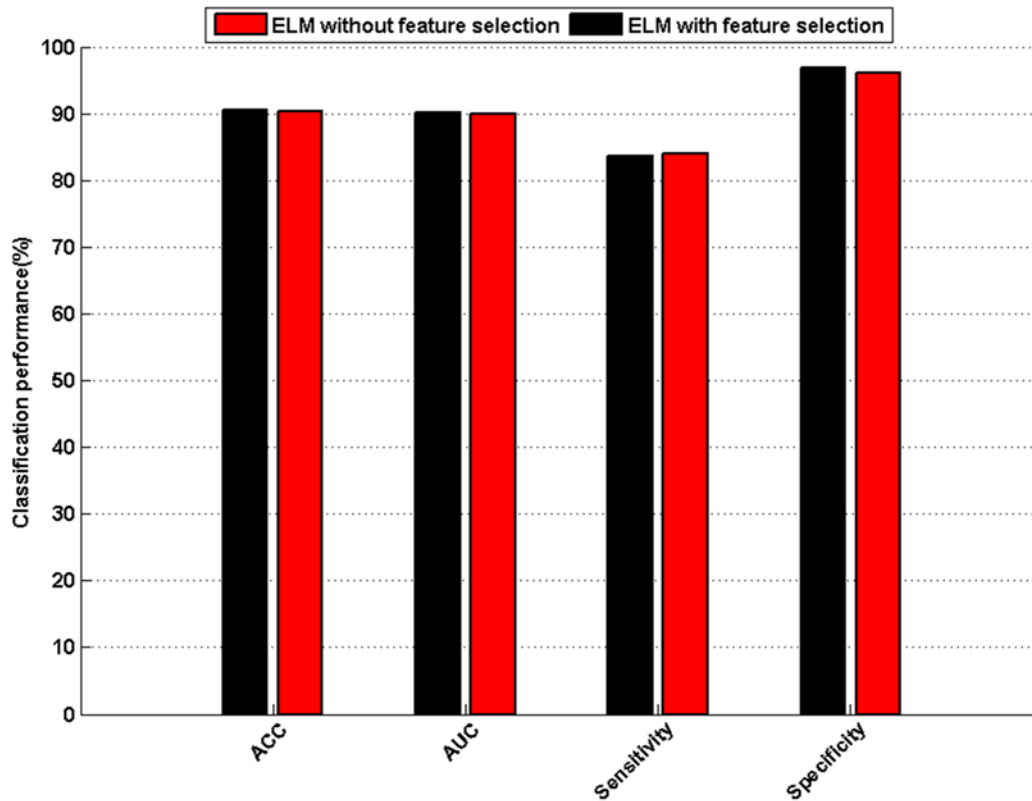


Fig 7. Comparison results of classification performance between ELM with and without feature selection.

doi:10.1371/journal.pone.0143003.g007

[32]. Obesity has been considered a systematic, low-level, chronic inflammation state that is heritable and predisposes the subject to many diseases [33]. Clearly detecting the overweight status could have great clinical significance.

In this study, routine blood tests were obtained and One-Way ANOVA analysis showed that there are significant differences between overweight and healthy subjects (p -value < 0.01) in their TG, GLU, LDL, HDL, CHO, ALT, AST, γ -GT, ALB, CR, BUN, AKP, TBIL, DBIL, UA, WBCU, PMC, AVM, RBC, HCT, AVLC, MCH, MCHC, AVE, PE, HB, AVN. Of these, TG, LDL, HDL and CHO belong to blood-lipid test indexes; ALT, AST, γ -GT, ALB, AKP, TBIL and DBIL belong to liver functions; CR, BUN and UA belong to renal functions; WBCU, PMC, AVM, RBC, HCT, AVLC, MCH, MCHC, AVE, PE, HB and AVN belong to blood routine tests. However, ANOVA analysis can only show differences, not correlation.

According to the results of the feature selection, the most important indexes that correlate with the overweight status are CR, HB, HCT, UA, RBC, HDL, ALT, TG and γ -GT, which is consistent with the results of the Spearman test analysis. These indexes have statistically

Table 6. The confusion matrix obtained by ELM with and without feature selection.

ELM with the whole feature set	Predicted overweight people	Predicted healthy controls
Overweight people	189	36
Healthy controls	10	241
ELM with the optimal feature set		
Overweight people	188	37
Healthy controls	8	243

doi:10.1371/journal.pone.0143003.t006

significant differences (p -value < 0.01). Among these blood parameters, CR and UA are renal function indexes. CR is a nitrogenous metabolic substance in muscles and blood. Uric acid is produced from the natural breakdown of cells and foods. Overweight subjects have fast metabolisms; they consume more calories per day than healthy subjects. Therefore, it is understandable that CR and UA were identified as two main features. HB is a protein in red blood cells; HCT is the volume percentage of red blood cells; RBC (erythrocytes) takes oxygen from the lungs and releases it into tissues. These are the three most important and relevant indexes for delivering oxygen. As overweight person consumes more calories and therefore requires more oxygen; this is consistent with elevated CR and UA. ALT and γ -GT are found mainly in the liver and are reliable indexes for liver damage. The liver is the metabolic factory in the human body; therefore, the obtained results in this study are reliable and have medical significance. The physiological state was different for overweight subjects; the renal function state, particularly the CR level, is a correlated parameter.

In this study, overweight and healthy subjects were classified according to BMI, which correlated with body fat. Body fat is the most common index used to assess the overweight status [34]. It also serves as a biomarker in hyperuricemia [35], and can predict the severity of coronary artery disease [36], nonalcoholic fatty liver disease [37], and the clustering of cardiometabolic risk factors in adolescents [38]. When BMI is measured at the base line, it was as effective as 2-h plasma glucose and fasting plasma glucose to predict diabetes [39]. Although BMI has been widely used as a diagnosis criterion for overweight subjects, it is not an accurate predicting indicator for some diseases. For instance, in middle-aged and older women, waist circumference is a more sensitive measurement of relative disease risk than is BMI [40]. The waist-to-height ratio showed better correlation with the severity of coronary artery disease than BMI [36]. Few studies examined in the relationship between BMI and blood biochemical changes in overweight subjects. Based on this study, selected blood indexes including HB, HCT, UA, RBC, HDL, ALT, TG and γ -GT can be used with BMI to assess the overweight status or predict some obesity-related diseases. Combined with the results of our findings, BMI will be a useful tool in the clinical practice.

Conclusion

This study attempted to determine an effective data-driven machine learning model for discriminating overweight from healthy controls using blood and biochemical indexes for the first time. In addition, correlating factors from the blood parameters were also investigated in overweight subjects by introducing the feature selection technique. A new machine learning paradigm, ELM, was explored and the classification performance was compared against the competitive SVM and BPNN in various evaluation metrics. Experimental results show that the ELM performs much more efficiently than the SVM and BPNN, and with higher recognition rates. Nine informative indexes that strongly correlate the overweight condition were identified with the feature selection; they include CR, HB, HCT, UA, RBC, HDL, ALT, TG and γ -GT.

Based on these discoveries, the proposed ELM-based approach for overweight detection in biomedical applications has promising potential. It provides a viable alternative solution to traditional overweight modeling tools by offering excellent predictive ability. However, it should be noted that only 225 overweight people were involved in the experiment; more data samples are needed to further verify the effectiveness of the proposed methodology.

Supporting Information

S1 Fig. Training accuracy surface of SVM with parameters obtained by grid search. The file lists the training accuracy surface of SVM with parameters obtained by grid search. (TIF)

S2 Fig. The hierarchical clustering heat map developed on the nine selected features. The file lists the hierarchical clustering heat map developed on the nine selected features. (TIF)

S1 Table. The random biases and input weights. The file lists the randomly produced biases and input weights produced by the ELM during training. (XLSX)

S2 Table. The classification performance of ELM obtained on the top ranked five, seven and nine features. The file lists the classification performance of ELM obtained on the top ranked five, seven and nine features. (DOCX)

Acknowledgments

Funding: This research is supported by the National Natural Science Foundation of China (NSFC) (61303113, 61133011, 61572226, 61572367, 61373053, 61272018, 81300311 and 61402337). This work is also supported by the key construction academic subject (medical innovation) of Zhejiang Province (11-CX26), Zhejiang Provincial Natural Science Foundation of China (R1110261, LQ13G010007, LQ13F020011, LY14F020035), and the young talents project of the First Affiliated Hospital of Wenzhou Medical University (qnyc 043), Jilin Province Natural Science Foundation under grants 20150101052JC, the Science and Technology Plan Project of Wenzhou, China (G20140048), the open project program of Key Laboratory of Symbolic Computation and Knowledge Engineering of Ministry of Education, Jilin University under Grant No. 93K172013K01, the Opening Project of Zhejiang Provincial Top Key Discipline of Pharmaceutical Sciences.

Author Contributions

Conceived and designed the experiments: HS YL XZ LH. Performed the experiments: HC LH. Analyzed the data: HC WL LH. Contributed reagents/materials/analysis tools: HC WL YL XZ LH. Wrote the paper: HC BY DL XZ LH.

References

1. Seo DC, Niu J. Trends in underweight and overweight/obesity prevalence in Chinese youth, 2004–2009. *International journal of behavioral medicine*. 2014; 21(4):682–90. doi: [10.1007/s12529-013-9322-1](https://doi.org/10.1007/s12529-013-9322-1) PMID: [23744048](https://pubmed.ncbi.nlm.nih.gov/23744048/).
2. Yan WL, Li XS, Wang Q, Huang YD, Zhang WG, Zhai XH, et al. Overweight, high blood pressure and impaired fasting glucose in Uyghur, Han, and Kazakh Chinese children and adolescents. *Ethnicity & health*. 2014;1–11. doi: [10.1080/13557858.2014.921894](https://doi.org/10.1080/13557858.2014.921894) PMID: [24904957](https://pubmed.ncbi.nlm.nih.gov/24904957/); PubMed Central PMCID: PMC4258184.
3. Vikoren LA, Nygard OK, Lied E, Rostrup E, Gudbrandsen OA. A randomised study on the effects of fish protein supplement on glucose tolerance, lipids and body composition in overweight adults. *The British journal of nutrition*. 2013; 109(4):648–57. doi: [10.1017/S0007114512001717](https://doi.org/10.1017/S0007114512001717) PMID: [22647247](https://pubmed.ncbi.nlm.nih.gov/22647247/).
4. Almoosawi S, Tsang C, Ostertag LM, Fyfe L, Al-Dujaili EA. Differential effect of polyphenol-rich dark chocolate on biomarkers of glucose metabolism and cardiovascular risk factors in healthy, overweight and obese subjects: a randomized clinical trial. *Food & function*. 2012; 3(10):1035–43. doi: [10.1039/c2fo30060e](https://doi.org/10.1039/c2fo30060e) PMID: [22796902](https://pubmed.ncbi.nlm.nih.gov/22796902/).
5. Kopelman PG. Obesity as a medical problem. *Nature*. 2000; 404(6778):635–43. PMID: [10766250](https://pubmed.ncbi.nlm.nih.gov/10766250/)
6. Hardalac F, Ozan AT, Barisci N, Ergun U, Serhatlioglu S, Guler I. The examination of the effects of obesity on a number of arteries and body mass index by using expert systems. *Journal of medical systems*. 2004; 28(2):129–42. doi: [10.1023/B:JOMS.0000023296.42481.a1](https://doi.org/10.1023/B:JOMS.0000023296.42481.a1) PMID: [MEDLINE:15195844](https://pubmed.ncbi.nlm.nih.gov/15195844/).

7. Zhang S, Tjortjis C, Zeng X, Qiao H, Buchan I, Keane J. Comparing data mining methods with logistic regression in childhood obesity prediction. *Information Systems Frontiers*. 2009; 11(4):449–60. doi: [10.1007/s10796-009-9157-0](https://doi.org/10.1007/s10796-009-9157-0) PMID: [WOS:000268574200009](https://pubmed.ncbi.nlm.nih.gov/2000268574200009/).
8. Ergun U. The Classification of Obesity Disease in Logistic Regression and Neural Network Methods. *Journal of Medical Systems*. 2009; 33(1):67–72. doi: [10.1007/s10916-008-9165-5](https://doi.org/10.1007/s10916-008-9165-5) PMID: [WOS:000262482900008](https://pubmed.ncbi.nlm.nih.gov/2000262482900008/).
9. Valavanis IK, Mouggiakakou SG, Grimaldi KA, Nikita KS. A multifactorial analysis of obesity as CVD risk factor: Use of neural network based methods in a nutrigenetics context. *BMC Bioinformatics*. 2010; 11:453–. doi: [10.1186/1471-2105-11-453](https://doi.org/10.1186/1471-2105-11-453) PMID: [PMC2941694](https://pubmed.ncbi.nlm.nih.gov/2002941694/).
10. Heydari ST, Ayatollahi SMT, Zare N. Comparison of Artificial Neural Networks with Logistic Regression for Detection of Obesity. *Journal of Medical Systems*. 2012; 36(4):2449–54. doi: [10.1007/s10916-011-9711-4](https://doi.org/10.1007/s10916-011-9711-4) PMID: [WOS:000306549000037](https://pubmed.ncbi.nlm.nih.gov/2000306549000037/).
11. Lee BJ, Kim KH, Ku B, Jang J-S, Kim JY. Prediction of body mass index status from voice signals based on machine learning for automated medical applications. *Artificial Intelligence in Medicine*. 2013; 58(1):51–61. doi: <http://dx.doi.org/10.1016/j.artmed.2013.02.001>. doi: [10.1016/j.artmed.2013.02.001](https://doi.org/10.1016/j.artmed.2013.02.001) PMID: [23453267](https://pubmed.ncbi.nlm.nih.gov/23453267/)
12. Huang G-B, Zhu Q-Y, Siew C-K. Extreme learning machine: Theory and applications. *Neurocomputing*. 2006; 70(1–3):489–501. doi: [10.1016/j.neucom.2005.12.126](https://doi.org/10.1016/j.neucom.2005.12.126)
13. Zhang R, Huang GB, Sundararajan N, Saratchandran P. Multicategory classification using an extreme learning machine for microarray gene expression cancer diagnosis. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*. 2007; 4(3):485–94. PMID: [17666768](https://pubmed.ncbi.nlm.nih.gov/17666768/)
14. Hu L, Hong G, Ma J, Wang X, Chen H. An efficient machine learning approach for diagnosis of paraquat-poisoned patients. *Computers in Biology and Medicine*. 59:116–24. doi: [10.1016/j.compbiomed.2015.02.003](https://doi.org/10.1016/j.compbiomed.2015.02.003) PMID: [25704654](https://pubmed.ncbi.nlm.nih.gov/25704654/)
15. Suresh S, Venkatesh Babu R, Kim HJ. No-reference image quality assessment using modified extreme learning machine classifier. *Applied Soft Computing*. 2009; 9(2):541–52. doi: [10.1016/j.asoc.2008.07.005](https://doi.org/10.1016/j.asoc.2008.07.005)
16. Mohammed AA, Minhas R, Jonathan Wu QM, Sid-Ahmed MA. Human face recognition based on multi-dimensional PCA and extreme learning machine. *Pattern Recognition*. 2011; 44(10–11):2588–97.
17. Pal M. Extreme-learning-machine-based land cover classification. *International Journal of Remote Sensing*. 2009; 30(14):3835–41.
18. Moreno R, Corona F, Lendasse A, Grana M, Galvao LS. Extreme learning machines for soybean classification in remote sensing hyperspectral images. *Neurocomputing*. 2014; 128:207–16. doi: [10.1016/j.neucom.2013.03.057](https://doi.org/10.1016/j.neucom.2013.03.057) PMID: [WOS:000331851700025](https://pubmed.ncbi.nlm.nih.gov/2000331851700025/).
19. Cortes C, Vapnik V. Support-vector networks. *Machine learning*. 1995; 20(3):273–97.
20. Hyska J, Mersini E, Mone I, Burazeri G. Prevalence and demographic correlates of overweight and obesity among children in a transitional southeastern European population. *Journal of community health*. 2014; 39(5):828–34. doi: [10.1007/s10900-014-9888-9](https://doi.org/10.1007/s10900-014-9888-9) PMID: [24878613](https://pubmed.ncbi.nlm.nih.gov/24878613/).
21. Duda RO, Hart PE, Stork DG. *Pattern classification*: John Wiley & Sons; 2012.
22. Huang G-B, Zhu Q-Y, Siew C-K. Extreme learning machine: a new learning scheme of feedforward neural networks. *IEEE International Joint Conference on Neural Networks*; 2004/072004. p. 985–90.
23. Huang GB, Babri HA. Upper bounds on the number of hidden neurons in feedforward networks with arbitrary bounded nonlinear activation functions. *Neural Networks, IEEE Transactions on*. 1998; 9(1):224–9.
24. Huang GB. Learning capability and storage capacity of two-hidden-layer feedforward networks. *Neural Networks, IEEE Transactions on*. 2003; 14(2):274–81.
25. Huang GB, Chen L, Siew CK. Universal approximation using incremental constructive feedforward networks with random hidden nodes. *Neural Networks, IEEE Transactions on*. 2006; 17(4):879–92.
26. Chang CC, Lin CJ. LIBSVM: a library for support vector machines. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>; 2001.
27. Salzberg SL. On comparing classifiers: Pitfalls to avoid and a recommended approach. *Data mining and knowledge discovery*. 1997; 1(3):317–28.
28. Fawcett T. An introduction to ROC analysis. *Pattern recognition letters*. 2006; 27(8):861–74.
29. Fawcett T. ROC graphs: Notes and practical considerations for researchers. *Machine Learning*. 2004; 31:1–38.
30. Liu T, Hu L, Ma C, Wang ZY, Chen HL. A fast approach for detection of erythemato-squamous diseases based on extreme learning machine with maximum relevance minimum redundancy feature selection.

International Journal of Systems Science. 2015; 46(5):919–31. doi: [10.1080/00207721.2013.801096](https://doi.org/10.1080/00207721.2013.801096) PMID: [WOS:000346069500014](https://pubmed.ncbi.nlm.nih.gov/25673413/).

31. Hsu CW, Chang CC, Lin CJ. A practical guide to support vector classification. Technical report, Department of Computer Science and Information Engineering, National Taiwan University, Taipei, 2003. available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>; 2003.
32. Ferenci T, Kovacs L, Almassy Z, Szilagyi L, Benyo B, Benyo Z. Differences in the laboratory parameters of obese and healthy Hungarian children and their use in automatic classification. Conference proceedings: Annual International Conference of the IEEE Engineering in Medicine and Biology Society IEEE Engineering in Medicine and Biology Society Annual Conference. 2010; 2010:3883–6. doi: [10.1109/IEMBS.2010.5627672](https://doi.org/10.1109/IEMBS.2010.5627672) PMID: [21097074](https://pubmed.ncbi.nlm.nih.gov/21097074/).
33. Locke AE, Kahali B, Berndt SI, Justice AE, Pers TH, Day FR, et al. Genetic studies of body mass index yield new insights for obesity biology. *Nature*. 2015; 518(7538):197–206. doi: [10.1038/nature14177](https://doi.org/10.1038/nature14177) PMID: [25673413](https://pubmed.ncbi.nlm.nih.gov/25673413/).
34. Evans EM, Rowe DA, Racette SB, Ross KM, McAuley E. Is the current BMI obesity classification appropriate for black and white postmenopausal women? *International journal of obesity*. 2006; 30(5):837–43. doi: [10.1038/sj.ijo.0803208](https://doi.org/10.1038/sj.ijo.0803208) PMID: [16418761](https://pubmed.ncbi.nlm.nih.gov/16418761/).
35. Lee MF, Liou TH, Wang W, Pan WH, Lee WJ, Hsu CT, et al. Gender, body mass index, and PPAR-gamma polymorphism are good indicators in hyperuricemia prediction for Han Chinese. *Genetic testing and molecular biomarkers*. 2013; 17(1):40–6. doi: [10.1089/gtmb.2012.0231](https://doi.org/10.1089/gtmb.2012.0231) PMID: [23237621](https://pubmed.ncbi.nlm.nih.gov/23237621/); PubMed Central PMCID: PMC3525899.
36. Sabah KM, Chowdhury AW, Khan HI, Hasan AT, Haque S, Ali S, et al. Body mass index and waist/height ratio for prediction of severity of coronary artery disease. *BMC research notes*. 2014; 7:246. doi: [10.1186/1756-0500-7-246](https://doi.org/10.1186/1756-0500-7-246) PMID: [24742250](https://pubmed.ncbi.nlm.nih.gov/24742250/); PubMed Central PMCID: PMC4000139.
37. Zheng RD, Chen ZR, Chen JN, Lu YH, Chen J. Role of Body Mass Index, Waist-to-Height and Waist-to-Hip Ratio in Prediction of Nonalcoholic Fatty Liver Disease. *Gastroenterology research and practice*. 2012; 2012:362147. doi: [10.1155/2012/362147](https://doi.org/10.1155/2012/362147) PMID: [22701476](https://pubmed.ncbi.nlm.nih.gov/22701476/); PubMed Central PMCID: PMC3369513.
38. Kong AP, Ko GT, Ozaki R, Wong GW, Tong PC, Chan JC. Metabolic syndrome by the new IDF criteria in Hong Kong Chinese adolescents and its prediction by using body mass index. *Acta paediatrica*. 2008; 97(12):1738–42. doi: [10.1111/j.1651-2227.2008.01056.x](https://doi.org/10.1111/j.1651-2227.2008.01056.x) PMID: [18945275](https://pubmed.ncbi.nlm.nih.gov/18945275/).
39. Rolandsson O, Hagg E, Nilsson M, Hallmans G, Mincheva-Nilsson L, Lernmark A. Prediction of diabetes with body mass index, oral glucose tolerance test and islet cell autoantibodies in a regional population. *Journal of internal medicine*. 2001; 249(4):279–88. PMID: [11298847](https://pubmed.ncbi.nlm.nih.gov/11298847/).
40. Van Pelt RE, Evans EM, Schechtman KB, Ehsani AA, Kohrt WM. Waist circumference vs body mass index for prediction of disease risk in postmenopausal women. *International journal of obesity and related metabolic disorders: journal of the International Association for the Study of Obesity*. 2001; 25(8):1183–8. doi: [10.1038/sj.ijo.0801640](https://doi.org/10.1038/sj.ijo.0801640) PMID: [11477503](https://pubmed.ncbi.nlm.nih.gov/11477503/).