Computers in Biology and Medicine 45 (2014) 34–42

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/cbm

A hybrid intelligent system for diagnosing microalbuminuria in type 2 diabetes patients without having to measure urinary albumin

Hamid R. Marateb [a], Marjan Mansourian [b,*], Elham Faghihimani [c], Masoud Amini [c], Dario Farina [d]

[a] Biomedical Engineering Department, Faculty of Engineering, University of Isfahan, Isfahan, Iran
[b] Department of Biostatistics and Epidemiology, School of Public Health, Isfahan University of Medical Sciences, HezarJerib St., 81745 Isfahan, Iran
[c] Isfahan Endocrine and Metabolism Research Center, Isfahan University of Medical Sciences, 81445-03500 Isfahan, Iran
[d] Department of NeuroRehabilitation Engineering, Bernstein Focus Neurotechnology Göttingen, Center for Computational Neuroscience, University Medical Center Göttingen, Georg-August University, 37075 Göttingen, Germany

ARTICLE INFO

Article history:
Received 28 September 2013
Accepted 18 November 2013

Keywords:
Classification
Diabetes type 2
Microalbuminuria
Expert-based system
Fuzzy rule induction
Particle swarm optimization
Statistical feature extraction

ABSTRACT

Microalbuminuria (MA) is an independent predictor of cardiovascular and renal disease, development of overt nephropathy, and cardiovascular mortality in patients with type 2 diabetes. Detecting MA is an important screening tool to identify people with high risk of cardiovascular and kidney disease. The gold standard to detect MA is measuring 24-h urine albumin excretion. A new method for MA diagnosis is presented in this manuscript which uses clinical parameters usually monitored in type 2 diabetic patients without the need of an additional measurement of urinary albumin. We designed an expert-based fuzzy MA classifier in which rule induction was performed by particle swarm optimization. A variety of classifiers was tested. Additionally, multiple logistic regression was used for statistical feature extraction. The significant features were age, diabetic duration, body mass index and HbA1C (the average level of blood sugar over the previous 3 months, which is routinely checked every 3 months for diabetic patients). The resulting classifier was tested on a sample size of 200 patients with type 2 diabetes in a cross-sectional study. The performance of the proposed classifier was assessed using (repeated) holdout and 10-fold cross-validation. The minimum sensitivity, specificity, precision and accuracy of the proposed fuzzy classifier system with feature extraction were 95%, 85%, 84% and 92%, respectively. The proposed hybrid intelligent system outperformed other tested classifiers and showed "almost perfect agreement" with the gold standard. This algorithm is a promising new tool for screening MA in type-2 diabetic patients.

1. Introduction

Diabetes mellitus is a group of metabolic diseases. A diabetic patient has high blood sugar, either because not enough insulin is produced by the pancreas, or because cells do not respond to the produced insulin [1]. Worldwide, 347 million people had diabetes in 2011 and by 2030 this number will increase to 552 million. Diabetes caused 4.6 million deaths in 2011 and is projected to be the 7th leading cause of death in 2030 according to the WHO. More than 80% of diabetes deaths occur in low- and middle-income countries [2]. Type 2 diabetes, the most common form of diabetes, results from the body's inefficient use of insulin and causes at least 90% of all cases of diabetes. Diabetic 2 patients have high risk of a variety of complications, basically because of complicated and interconnected mechanisms such as hyperglycemia, insulin-resistance, and accelerated atherogenesis. [3]. Life threatening cardio-cerebrovascular disease such as coronaropathy, stroke and heart failure are also associated with type 2 diabetes.

Microalbuminuria (MA) is one of the first clinical indicators of microvascular damage in diabetes [4]. MA is defined as a persistent elevation of albumin in the urine of >30 to <300 mg/d (>20 to <200 μg/min). The acceptable amount of albumin in the urine is less than 30 mg/d; values above 300 mg/d (200 μg/min) indicate overt proteinuria (macroalbuminuria). The underlying association between MA and the development of diabetic complications has been well established in the literature [5]. In fact, MA is a well-known predictor of renal disease and consequent progress of overt diabetic nephropathy in patients with type 2 diabetes [6,7]. It has also been shown that MA predicts silent myocardial ischemia, polyvascular diseases, and increased (cardiovascular)

* Corresponding author. Tel.: +98 311 7922651; fax: +98 311 668864.
E-mail addresses: h.marateb@eng.ui.ac.ir (H.R. Marateb), j_mansourian@hlth.mui.ac.ir (M. Mansourian), efimani@yahoo.com (E. Faghihimani), m_amini@med.mui.ac.ir (M. Amini), dario.farina@bccn.uni-goettingen.de (D. Farina).

0010-4825/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.compbiomed.2013.11.006

mortality in type 2 diabetes patients. Considering the high incidence of MA (approximately 40%) in patients with type 2 diabetes [8] and its significant association with cardiovascular and renal events, its screening and intervention measures are very important.

Studies have shown that many disorders, such as hyperglycemia, hypertension, and obesity in diabetic patients are associated with an increased risk of MA [9]. Also, several studies have revealed that the risk of MA depends on multiple factors, including age, gender, BMI (body mass index), DD (diabetic duration), BP (blood pressure), FBS (fasting blood sugar), HbA1c (the average percentage level of blood glucose over the previous 3 months), Bs2hpp (2 h post-prandial blood glucose), CHOL (total cholesterol), LDL (low-density lipoprotein), HDL (high-density lipoprotein), and TG (triglyceride) [10]. Previous studies on the relationship between various risk factors and MA have provided controversial results. In the literature, linear regression methods were usually used for analyzing the risk factors of MA. They include either univariate methods (the correction between the factor and the outcome) or multivariate ones (multiple linear regressions). Univariate methods do not consider the interaction between different factors. Multivariate techniques, on the other hand, depend on several assumptions, including the linearity in the underlying system or the normality of the response variable, assumptions which are often not valid for real data sets. Many physiological systems are highly nonlinear [11], so that linear analysis is inappropriate. Advanced data-mining methods, on the other hand, can be used to extract high-level information from complex biological systems without such troublesome pre-assumptions.

MA can be diagnosed by 24-h urine collection (gold standard). However, more convenient detection methods have been proposed in the literature [12]. The present study investigates whether it is possible to use advanced classification methods to indicate the occurrence of MA based *solely* on the routine screening in diabetic patients [13], without a separate urinary albumin concentration measurement test. Considering the rather high prevalence of chronic kidney disease in diabetic patients in Iran (14%–26%) [14], this will facilitate MA diagnosis, thus preventing later kidney complications. Moreover, this study could reveal which are the most significant factors for predicting MA, thus simplifying the diagnostic procedure.

The rest of the paper is organized as follows: in the next section, information about the subjects and the classification methods used in this study is presented. Section 3 provides the results of the classification methods and assesses their performance. Statistical feature selection is used to exclude irrelevant features. Finally, the conclusions are summarized in Section 4.

2. Materials and methods

2.1. Patients

Participants attended screening visits and a follow-up visit at the Endocrine and Metabolism Research Center, Isfahan University of Medical Sciences, where they underwent standardized interviews, physical assessment, and laboratory testing. We studied all the diabetic 2 patients who attended the above-mentioned center in 2012 whose clinical information was complete. The data were collected by trained technicians. The duration of diabetes was determined by subtracting the age at diabetes diagnosis from the current age. MA was defined as urinary albumin between 30 and 300 mg/day. Lipid profile (HDL, LDL, TG and CHOL) was measured after a 10- to 12-h fast. Our study was performed on 200 patients (130 males and 70 females) with type 2 (insulin-Independent) diabetes mellitus. All patients received conventional treatment. Also semiannual determinations of HbA1c, Bs2hpp, BP, BMI, and FBS values were recorded for participants. All subjects gave informed consent to the experimental procedure. The experimental protocol was approved by the Isfahan University of Medical Sciences Panel on Medical Human Subjects and conformed to the Declaration of Helsinki.

2.2. Statistical analysis

All statistical analyses and calculations were performed using the SPSS statistical package, version 18.0 (SPSS Inc., Chicago, IL, USA). Descriptive analyses were used to characterize the participants by socio-demographic and clinical factors. Continuous data were presented as mean $\pm$ SD and as proportions for categorical variables. Prevalence of MA was expressed as percentages. A two-sided P-value of <0.05 was considered statistically significant. Finally, classification was performed using Matlab and Statistics Toolbox Release 2011a (The MathWorks, Inc., Natick, Massachusetts, USA).

2.3. Classification methods

The following input features were used in this study: Age, Gender, BMI, DD, systolic BP, FBS, HbA1c, Bs2hpp, CHOL, LDL, HDL, and TG. BP was then converted to a nominal variable indicating the occurrence of hypertension. This was done because BP has been usually used to indicate hypertension as a risk factor in the literature. Two nominal variables (gender, BP) received special treatment when necessary. The outcome was classified as Normoalbuminuria (urinary albumin less than 30 mg/day) or MA (urinary albumin between 30 and 300 mg/day).

Machine learning methods applied to a variety of medical domains can be classified into the following major groups [15,16]: inductive learning of symbolic/qualitative rules (such as decision trees and neuro-fuzzy classifiers), statistical or pattern-recognition methods (discriminate analysis, Bayesian classifiers, support vector machines), and artificial neural networks. Among multivariate methods, linear and quadratic discriminant analysis (LDA and QDA), support vector machine (SVM), naïve Bayesian classifier (NBC), bagged decision trees (BDT), and adaptive neuro-fuzzy interference system (ANFIS) were chosen [16–18]. Also, we proposed an expert-based type-1 Mamdani fuzzy inference system, in which the rule fusion was performed using discrete particle swarm optimization (D-PSO). A statistical feature selection (FS) was also used to reduce the feature space. Its effect on the examined classifiers was studied. In the next section, the classification methods used in this study are introduced.

2.3.1. The proposed hybrid intelligent system (HIS)

Fuzzy classifiers (FCs) have been widely used in computer-aided diagnosis systems [19,20]. FCs assume an overlapping boundary between neighboring classes which is practical in many applications, providing a simple representation of the complex feature space.

Briefly, fuzzy *if-then* rules in a two-class classifier FIS are defined as follows:

$$\text{if } x_1 \text{ is } A_1^j \text{ and } x_2 \text{ is } A_2^j \text{ and } \ldots \text{ and } x_n \text{ is } A_n^j \text{ then } y \text{ is } \omega_i; \quad i=1,\ 2\ (w_j) \tag{1}$$

where, A_i^j and w_j are the ith input linguistic term and the weight (importance) of the rule index j ($j=1,\ldots, m$), and $[x_1,\ldots,x_n]$ is the feature vector (the clinical recordings for a subject) that is classified to class ω_i. Similarly, it is possible to change the rule consequent to a probability in which the input data belongs to one