

“© 2013 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.”

# Hypoglycaemia Detection using Fuzzy Inference System with Multi-objective Double Wavelet Mutation Differential Evolution

J.C.Y. Lai<sup>a</sup>, F.H.F. Leung<sup>b</sup>, S.H. Ling<sup>c</sup> and H.T. Nguyen<sup>d</sup>

<sup>a</sup>Centre for Signal Processing, Dept. of Electronic and Information Engg., The Hong Kong Polytechnic University, Hung Hom, Hong Kong. E-mail: 08900438r@polyu.edu.hk

<sup>b</sup>Centre for Signal Processing, Dept. of Electronic and Information Engg., The Hong Kong Polytechnic University, Hung Hom, Hong Kong. E-mail: enfrank@inet.polyu.edu.hk

<sup>c</sup>Centre for Health Technologies, Faculty of Engineering and IT, University of Technology Sydney, NSW2007, Australia. E-mail: Steve.Ling@uts.edu.au

<sup>d</sup>Centre for Health Technologies, Faculty of Engineering and IT, University of Technology Sydney, NSW2007, Australia. E-mail: Hung.Nguyen@uts.edu.au

## ABSTRACT

In this paper, a fuzzy inference system (FIS) is developed to recognize hypoglycaemic episodes. Hypoglycaemia (low blood glucose level) is a common and serious side effect of insulin therapy in patients with diabetes. We measure physiological parameters continuously to provide hypoglycaemia detection for Type 1 diabetes mellitus (T1DM) patients. The FIS captures the relationship between the inputs of heart rate (HR), corrected QT interval of the electrocardiogram (ECG) signal ( $QT_c$ ), change of HR, change of  $QT_c$  and the output of hypoglycaemic episodes to perform the classification. A differential evolution with double wavelet mutation operations (DWM-DE) is introduced to optimize the FIS parameters that govern the membership functions and fuzzy rules. DWM-DE is an improved differential evolution that incorporates two wavelet-based operations to enhance the optimization performance. To prevent the phenomenon of overtraining (over-fitting), a validation approach is proposed. Moreover, in this problem, two targets of sensitivity and specificity should be met in order to achieve good performance. As a result, a multi-objective optimization using DWM-DE is introduced to perform the training of the FIS. Experiments using data of 15 children with T1DM (569 data points) are studied. The

real data are randomly organized into a training set with 5 patients (199 data points), a validation set with 5 patients (177 data points) and a testing set with 5 patients (193 data points). The result shows that the proposed fuzzy inference system tuned by the multi-objective DWM-DE can offer good performance of classification.

Keywords: Differential evolution, fuzzy inference system, hypoglycaemia, multi-objective optimization.

## 1. INTRODUCTION

Hypoglycaemia is the medical term for a body state produced by a low level of blood glucose. It is a result of the mismatch between the action of insulin, the ingestion of food and the energy expenditure. Hypoglycaemia is less common in non-diabetic persons, but can occur at any age [1]. Its causes, among others, can be excessive insulin produced in the body, inborn errors, medications and poisons, alcohol, hormone deficiencies, prolonged starvation, alterations of metabolism associated with infection, and organ failure [2]. It has been discussed that diabetic patients, especially those who have been treated with insulin, are at risk of developing hypoglycaemia. Most surveys reported that the tighter the glycaemia control and the younger the patient, the greater frequency of both mild and severe hypoglycaemia would occur [3]. The level of blood glucose to define hypoglycaemia may be different for different people, in different circumstances, and for different purposes. Most healthy adults maintain fasting glucose levels above 70 mg/dL (3.9 mmol/L), and develop symptoms of hypoglycaemia when the glucose level falls below 60 mg/dL (3.3 mmol/L) [2]. It has been reported that severe hypoglycaemic episodes are defined as documented blood glucose levels being below 50 mg/dL (2.8 mmol/L), and the

patients are advised to take necessary medical treatment. In some cases, it is treated by injection or infusion of glucagon. In fact, the brain and nervous system need to maintain a certain level of glucose in order to function. Two typical symptoms of hypoglycaemia arise from the activation of the autonomic central nervous systems (autonomic symptoms) and reduced cerebral glucose consumption (neuroglycopenic symptoms) [5]. Autonomic symptoms such as headache, extreme hunger, blurry or double vision, fatigue, weakness and sweating are activated before neuroglycopenic symptoms. Thus, the presence of hypoglycaemia can be evident when autonomic symptoms occur. Neuroglycopenic symptoms such as confusion, seizures, and loss of consciousness (coma) arise when insufficient glucose is supplied to the brain [6].

Many cases of hypoglycaemia are presented where the patients are unaware of their symptoms. Nocturnal hypoglycaemia is particularly dangerous because it may obscure autonomic counter-regulatory responses, so that any initially mild episodes may become severe. It was estimated that 50% of all severe episodes occur at night time. Deficient glucose counter-regulation may also lead to severe hypoglycaemia even with modest insulin elevation. Regulation of nocturnal hypoglycaemia is further complicated by the dawn phenomenon [7]. This is a consequence of nocturnal changes in insulin sensitivity secondary to growth hormone secretion: a decrease in insulin requirement approximately between midnight and 5 am followed by an increase in requirement between 5 am and 8 am. Thus, hypoglycaemia is one of the complications of diabetes most feared by patients. Constructing a model for the episodes of hypoglycaemia with respect to some physiological parameters are very important for the patients to perform real-time monitoring of their blood glucose level [8] [11] [12].

In this paper, a fuzzy inference system (FIS) is developed to model the relationship between four physiological parameters and the episodes of hypoglycaemia. The physiological parameters are the heart rate (HR), the change of HR with time ( $\Delta HR$ ), the corrected QT interval ( $QT_c$ ) of the electrocardiogram (ECG) signal and the change of  $QT_c$  with time ( $\Delta QT_c$ ). Fuzzy inference is a process of making decisions by using fuzzy logic and fuzzy rules. The use of fuzzy logic in rule-based systems has been a success, with reported applications in climate control, medicine, relational database, scheduling, etc. The process of fuzzy inference involves the definitions of membership functions, logical operations and fuzzy if-then rules. The membership functions formulate the description of linguistic terms for the inputs and outputs. The logical operations and fuzzy if-then rules can easily be derived based on human knowledge. FIS uses linguistic control rules obtained from experienced human experts to perform decision-making. Before the FIS can work as a classifier of hypoglycaemia for Type 1 diabetes mellitus (T1DM) patients, it has to be trained by some data set with known class labels. It is a kind of supervised learning. The major objective of the training process is to determine the best parameter values for the FIS's rules and membership functions. The training can be regarded as an optimization process. Traditional optimization methods like the least square algorithm and gradient descent methods have the potential problem of trapping in some local optima of the solution space. Hence, evolutionary optimization algorithms are being considered in this paper. An improved differential evolution with double wavelet mutation operations (DWM-DE) is proposed as the optimization method for the training of the FIS.

Differential evolution (DE) has been well accepted as a powerful algorithm for handling optimization problems during the last decade. Proposed by Storn and Price,

DE is a population based stochastic optimization algorithm that searches the solution space by using the weighted difference between two population vectors to determine a third vector [9] [17] [18]. Owing to the population-based strategy, DE is less possibly getting trapped in some locally optimal solution. To enhance the performance of DE, we propose the DWM-DE. By applying the wavelet operation in the DE's mutation and crossover, we can have the solution space more widely explored in the early stage of the search. We are also more likely to obtain a fine-tuned global solution in the later stage of the search by setting a smaller searching space. The wavelets properties enable us to improve the performance of DE in terms of convergence speed, solution quality and solution stability statistically.

To measure the performance of the biomedical classification, sensitivity and specificity are introduced. The sensitivity measures the proportion of actual positives that are correctly identified; and the specificity measures the proportion of actual negatives that are correctly identified. In this application, the sensitivity should be higher than 70% and the specificity should be higher than 50% in order to obtain a reliable classification. The sensitivity of the hypoglycaemia detection system is more significant than the specificity because sensitivity concerns sick people. It is important to assess the abnormal condition for sick people accurately, due to the necessity for instant treatment. Based on these requirements, there are two objectives needed to be achieved in the training process. As a result, a multi-objective optimization approach using DWM-DE should be employed to realize the training process. Besides, overtraining is another key problem that affects the classification performance when constructing the FIS model. Overtraining refers to the reduction of the generalization ability that can occur as a system is trained. A lot of research has been done on reducing the overtraining (over-fitting) when using evolutionary

computation algorithms for optimizing system models [20][21]. In this paper, a validation strategy is proposed to reduce the risk of overtraining. This strategy is embedded in the training operation and the formulation of the fitness functions in DWM-DE. The details will be given in Section 2.

The organization of this paper is as follows: In Section 2, the details of the development of the FIS and the DWM-DE algorithm are presented. The experiment results for detecting nocturnal hypoglycaemic episodes in T1DM patients are discussed in Section 3. A conclusion is drawn in Section 4.

## 2. THE FUZZY INFERENCE SYSTEM WITH DWM-DE

To realize the detection of hypoglycaemic episodes for the T1DM cases, an FIS tuned by DWM-DE is proposed. A block diagram of the FIS is shown in Fig 1. It is a 4 inputs and 1 output system. The physiological inputs are the heart rate (HR), the corrected QT interval of the electrocardiogram signal ( $QT_c$ ), the change of heart rate ( $\Delta HR$ ) and the change of  $QT_c$  ( $\Delta QT_c$ ). The output is the binary state of hypoglycaemia ( $h$ ), which takes the value of true (+1) or false (-1). The major role of the FIS is to model the relationship of HR,  $QT_c$ ,  $\Delta HR$  and  $\Delta QT_c$  to  $h$  so as to perform classification.

Two of the inputs of the FIS are related to the ECG signal. The ECG signal being investigated in this paper involves the parameters in the depolarization and repolarisation stages of electrocardiography (Fig. 2). The concerned points are the Q points, R peak, T wave peak and the T wave end. The QT interval is between the Q point and the T wave peak. The  $QT_c$  signal is obtained by  $QT/(RR)$  where RR is the interval between two consecutive R peaks. The heart rate is given by  $60/RR$ .

## 2.1. Fuzzy Inference System

The physiological parameters (HR, QT<sub>c</sub>, ΔHR and ΔQT<sub>c</sub>) are represented as linguistic variables governed by membership functions that support fuzzy detections. The FIS consists of four major parts: fuzzification, fuzzy inferencing, fuzzy rules and defuzzification.

On doing fuzzification, the inputs are mapped to different membership functions by using fuzzy sets. In this paper, bell-shaped fuzzy membership functions are used and defined as follows:

$$f_m(x(t), \sigma_m, c_m) = e^{-\frac{(x-c_m)^2}{2\sigma_m^2}} \quad (1)$$

where  $x(t)$  is the non-fuzzy input,  $m=1, 2, \dots, m_f$ ;  $m_f$  denotes the number of membership functions;  $\sigma_m$  and  $c_m$  are the standard deviation and the mean value of the membership function respectively. Under the current problem, the membership functions for the four inputs are given by:

$$f_{HR,m}(HR(t), \sigma_{HR,m}, c_{HR,m}) = e^{-\frac{(HR(t)-c_{HR,m})^2}{2(\sigma_{HR,m})^2}} \quad (2)$$

$$f_{QT_c,m}(QT_c(t), \sigma_{QT_c,m}, c_{QT_c,m}) = e^{-\frac{(QT_c(t)-c_{QT_c,m})^2}{2(\sigma_{QT_c,m})^2}} \quad (3)$$

$$f_{\Delta HR,m}(\Delta HR(t), \sigma_{\Delta HR,m}, c_{\Delta HR,m}) = e^{-\frac{(\Delta HR(t)-c_{\Delta HR,m})^2}{2(\sigma_{\Delta HR,m})^2}} \quad (4)$$

$$f_{\Delta QT_c,m}(\Delta QT_c(t), \sigma_{\Delta QT_c,m}, c_{\Delta QT_c,m}) = e^{-\frac{(\Delta QT_c(t)-c_{\Delta QT_c,m})^2}{2(\sigma_{\Delta QT_c,m})^2}} \quad (5)$$



In the inference process, the fuzzified inputs are processed with respect to a rule base, which generates the fuzzified output. The rule base contains a set of fuzzy if-then rules of the following format:

Rule $\tau$ . <b>IF</b>	$HR(t)$	is	$f_{HR,m}(HR(t), \sigma_{HR,m}, c_{HR,m})$
<b>AND</b>	$QT_c(t)$	is	$f_{QT_c,m}(QT_c(t), \sigma_{QT_c,m}, c_{QT_c,m})$
<b>AND</b>	$\Delta HR(t)$	is	$f_{\Delta HR,m}(\Delta HR(t), \sigma_{\Delta HR,m}, c_{\Delta HR,m})$
<b>AND</b>	$\Delta QT_c(t)$	is	$f_{\Delta QT_c,m}(\Delta QT_c(t), \sigma_{\Delta QT_c,m}, c_{\Delta QT_c,m})$ ,
<b>THEN</b>	$y(t)$	is	$w_\tau$

where  $\tau$  is the rule number,  $\tau=1,2, \dots, n_r$ ;  $n_r$  is the number of rules in the fuzzy rule base. The value of  $n_r$  is defined by the following equation:

$$n_r = (m_f)^n \quad (6)$$

where  $n = 4$  is the number of fuzzy inputs for the FIS. After fuzzifying the input signals, aggregation is then realized to each rule to represent the output as a single fuzzy set. The aggregation output for each rule is defined as follows:

$$o_\tau = f_{HR,m}(HR(t), \sigma_{HR,m}, c_{HR,m}) \times f_{QT_c,m}(QT_c(t), \sigma_{QT_c,m}, c_{QT_c,m}) \times f_{\Delta HR,m}(\Delta HR(t), \sigma_{\Delta HR,m}, c_{\Delta HR,m}) \times f_{\Delta QT_c,m}(\Delta QT_c(t), \sigma_{\Delta QT_c,m}, c_{\Delta QT_c,m}) \quad (7)$$

During the inference process, we map the fuzzy set outputs of the rules to a crisp output by performing a process called defuzzification. In practice, defuzzification is a process of generating quantifiable results from the fuzzy sets and rules. In this paper, the output of defuzzification is realized as:

$$y(t) = \frac{\sum_{\tau=1}^{n_r} o_\tau w_\tau}{\sum_{\tau=1}^{n_r} o_\tau} \quad (8)$$

where  $w_\tau \in \{-1, 1\}$  is the fuzzy singleton to be determined. If the output  $y(t)$  is positive, it means that hypoglycaemia is present. ( $h(t)=+1$ ).

## 2.2. Differential Evolution with Double Wavelet Mutations

To implement the optimization process using Differential Evolution, a population in the solution space should be randomly generated. The population of solution vectors is then successively updated and swapped until the population converges to the optimum within the solution space. The pseudo code for the standard DE (SDE) process is shown in Fig. 3. In this paper, a DE with double wavelet mutation (DWM-DE) is proposed to tune the system parameters of the FIS. The pseudo code of the DWM-DE is shown in Fig. 4. The details of both the SDE and the DWM-DE are given as follows.

### 2.2.1. Standard Differential Evolution (SDE)

DE attempts to maintain a population of  $N_p$  vectors for each generation of evolution, with each vector contains  $D$  elements. Let  $P_{x,g}$  be the population of the current generation  $g$ , and  $\mathbf{x}_{i,g}$  be the  $i$ -th vector in this population:

$$\begin{aligned} P_{x,g} &= (\mathbf{x}_{i,g}), i = 0, 1, \dots, N_p - 1; g = 0, 1, \dots, g_{\max} \\ \mathbf{x}_{i,g} &= (x_{j,i,g}), j = 0, 1, \dots, D - 1. \end{aligned} \quad (9)$$

where  $g_{\max}$  is the maximum generation number. Before the population can be initialized over the solution space, the boundary of the searching space should be specified. The population should be uniformly and randomly distributed in the searching space. Once initialized, DE creates a mutated vector  $\mathbf{v}_{i,g}$  for each target

vector  $\mathbf{x}_{i,g}$  by using the mutation operation. This operation adds a scaled, randomly sampled, vector difference to  $\mathbf{x}_{i,g}$  to form a third vector. The mutated vector is therefore realized by the following equation:

$$\mathbf{v}_{i,g} = \mathbf{x}_{i,g} + F \cdot (\mathbf{x}_{r_1,g} - \mathbf{x}_{r_2,g}) \quad (10)$$

where  $F$  is the scaling factor;  $r_1$  and  $r_2$  are two different integers, which are randomly generated from  $\{0, 1, \dots, N_p-1\}$ . To complement the differential mutation search strategy and increase the diversity of the perturbed vectors, DE employs a method called uniform crossover for all the mutated vectors. Each vector element pair  $x_{j,i,g}$  and  $v_{j,i,g}$  generates a new trial vector element  $u_{j,i,g}$ , which is realized by the following equation:

$$\mathbf{u}_{i,g} = (u_{j,i,g}) = \begin{cases} (v_{j,i,g}) & \text{if } (rand_j(0,1) \leq C_r) \\ (x_{j,i,g}) & \text{otherwise.} \end{cases} \quad (11)$$

where  $C_r \in [0, 1]$  is called the crossover rate, which is a user-defined value that controls the fraction of parameters copied from the mutant.  $rand_j(0,1)$  generates a random value between 0 and 1 for the  $j$ -th parameter. The algorithm also ensures  $u_{j,i,g}$  gets at least one parameter value as  $x_{j,i,g}$ . Then the population is updated by comparing each trial vector  $\mathbf{u}_{i,g}$  to the corresponding target vector  $\mathbf{x}_{i,g}$ . If the fitness function value of the trial vector is smaller than that of the target vector, replace the target vector with the trial vector in the next generation; otherwise the target vector retains its place in the population for another generation. The selection operation is therefore realized by the following equation:

$$\mathbf{x}_{i,g+1} = \begin{cases} \mathbf{u}_{i,g} & \text{if } f(\mathbf{u}_{i,g}) \leq f(\mathbf{x}_{i,g}) \\ \mathbf{x}_{i,g} & \text{otherwise.} \end{cases} \quad (12)$$

where  $f(\cdot)$  is the fitness function. Because of this selection operation, DE is expected to have high optimization ability. When the condition to stop further evolution is

satisfied; for example, a preset maximum number of iteration has been reached, the algorithm ends with the best solution as the final solution.

### 2.2.2. Differential Evolution with Double Wavelet Mutation (DWM-DE)

In the SDE mutation operation, the value of  $F$  in (10) is a fixed value within the range of  $[0, 1]$  depending on the kind of application. The choice of this value relies very much on experience or expert knowledge. Yet, a fixed value of  $F$  takes no advantage of the benefit brought by the evolution. We propose the value of  $F$  to diminish with the increase of the number of iteration. Moreover, for some complex optimization problem such as finding the minimum point of a multimodal function with many local minima, a large number of iterations for solving the problem are required. It reduces the efficiency of the SDE. This leads to the proposed DWM-DE in which the value of  $F$  is determined by a wavelet function. The extent of different movements of the trial vectors will then be increased. More ‘random’ directions for the exploration would be generated during the mutation operation. Moreover, in the crossover operation, we proposed a second wavelet mutation that varies the searching space based on the wavelet function. As the wavelet function output is set to be inversely proportional to the number of iterations; when the searching population is approaching the optimal solution, the effect of the double wavelet mutations will be decreasing until the DE ends eventually. By adopting this method, the effort on searching and evaluating those local optima, which could be far away from the global optimum, in the later iteration is reduced. The total number of iterations should also decrease. Thanks to the property of the wavelet function, the solution stability is enhanced in a statistical sense, i.e. the performance of this hybrid DE on converging

to the optimal point is relatively stable despite the presence of many random factors during the evolution.

### 2.2.3. Double Wavelet Mutation

#### 2.2.3.1. Wavelet Function

Certain seismic signals can be modelled by combining translations and dilations of an oscillatory function called a “wavelet”. A continuous-time function  $\psi(x)$  is called a “mother wavelet” or “wavelet” if it satisfies the following properties:

Property 1:

$$\int_{-\infty}^{+\infty} \psi(x) dx = 0 \quad (13)$$

In other words, the total positive momentum of  $\psi(x)$  is equal to the total negative momentum of  $\psi(x)$ .

Property 2:

$$\int_{-\infty}^{+\infty} |\psi(x)|^2 dx < \infty \quad (14)$$

which means most of the energy in  $\psi(x)$  is confined to a finite domain and bounded.

The Morlet wavelet [16], as shown in Fig. 7, is an example of a mother wavelet:

$$\psi(x) = e^{-x^2/2} \cos(5x) \quad (15)$$

The Morlet wavelet integrates to zero (Property 1). Over 99% of the total energy of the function is contained in the interval of  $-2.5 < x < 2.5$  (Property 2). In order to control the magnitude of  $\psi(x)$ , a function  $\psi_a(x)$  is defined as follows.

$$\psi_a(x) = \frac{1}{\sqrt{a}} \psi\left(\frac{x}{a}\right) \quad (16)$$

where  $a$  is the dilation parameter.

It follows that  $\psi_a(x)$  is an amplitude-scaled version of  $\psi(x)$ . Fig. 8 shows different dilations of the Morlet wavelet. The amplitude of  $\psi_a(x)$  will be scaled down as the dilation parameter  $a$  increases. This property is used to do the mutation operation in order to enhance the searching performance.

### 2.2.3.2. Operation of DE with wavelet controlled mutation

The vectors in the population are mutated based on a proposed wavelet mutation (WM) operation, which exhibits a fine-tuning property. First, modify the mutation operation (10) as follows.

$$\mathbf{v}_{i,g} = \mathbf{x}_{i,g} + F \cdot \left( \mathbf{x}_{r_1,g} - \mathbf{x}_{r_2,g} \right), \quad (17)$$

where

$$F = \psi_a(\varphi), \quad (18)$$

$$F = \frac{1}{\sqrt{a}} \psi\left(\frac{\varphi}{a}\right). \quad (19)$$

By using the Morlet wavelet in (15) as the mother wavelet,

$$F = \frac{1}{\sqrt{a}} e^{-\left(\frac{\varphi}{a}\right)^2 / 2} \cos\left(5\left(\frac{\varphi}{a}\right)\right), \quad (20)$$

Referring to Property 1 of (13), the total positive momentum of the mother wavelet is equal to its total negative momentum. Then, the sum of the positive  $F$  is approximately equal to the sum of the negative  $F$  when the number of samples is large and  $\varphi$  is randomly generated, i.e.

$$\frac{1}{N} \sum_N F \approx 0 \text{ for } N \rightarrow \infty, \quad (21)$$

where  $N$  is the number of samples. Hence, the overall positive mutation and the overall negative mutation throughout the evolution are nearly the same in a statistical sense. This property gives better solution stability such that a smaller standard deviation of the solution values upon many trials can be reached. As over 99% of the total energy of the mother wavelet function is contained in the interval  $[-2.5, 2.5]$ ,  $\varphi$  can be generated from  $[-2.5a, 2.5a]$  randomly. The value of the dilation parameter  $a$  is set to vary with the value of  $t/T$  in order to meet the fine-tuning purpose, where  $T$  is the total number of iteration and  $t$  is the current number of iteration. In order to perform a local search when  $t$  is large, the value of  $a$  should increase as  $t/T$  increases so as to reduce the significance of the mutation. Hence, a monotonic increasing function governing  $a$  and  $t/T$  is proposed as follows.

$$a = e^{-\ln(\lambda) \times \left(1 - \frac{t}{T}\right)^{\zeta_{wm}} + \ln(\lambda)} \quad (22)$$

where  $\zeta_{wm}$  is the shape parameter of the monotonic increasing function,  $\lambda$  is the upper limit of the parameter  $a$ .

The effects of the various values of the shape parameter  $\zeta_{wm}$  to  $a$  with respect to  $t/T$  are shown in Fig. 9. In this figure,  $\lambda$  is set as 10000. Thus, the value of  $a$  is between 1 and 10000. Referring to (20), the maximum value of  $F$  is 1 when the random number of  $\varphi=0$  and  $a=1$  (at  $t/T = 0$ ). Then referring to (17), the vector  $\mathbf{v}_{i,g}$  has a large degree of mutation. It ensures that a large search space for the mutated vector is given at the early stage of evolution. When the value  $t/T$  is near to 1, the value of  $a$  is so large that the maximum value of  $F$  will become very small. For example, at  $t/T = 0.9$  and  $\zeta_{wm} = 1$ ,  $a = 400$ ; if the random value of  $\varphi$  is zero, the value

of  $F$  will be 0.0158. A smaller searching space for the mutated vector is then given for fine-tuning.

### 2.2.3.3. Operation of DE crossover with wavelet mutation

The crossover operation of (11) is done with respect to the elements of the trial vector (after mutation) in DE. In DWM-DE, a second-stage wavelet mutation is embedded in the crossover operation. It is realized by adding a second wavelet mutation following the original crossover operation. The details are as follows. The crossover after the first mutation takes place according to (11). Let  $\mathbf{u}_{i,g} = (u_{0,i,g}, u_{1,i,g}, \dots, u_{D-1,i,g})$  (where  $g$  is the current generation number and  $D$  is the number of elements in the vector) be the  $i$ -th vector after crossover for the second wavelet mutation. The value of the element  $u_{j,i,g}$  is inside the vector element's boundary  $[para_{\min}^j, para_{\max}^j]$ . The mutated crossover vector is given by

$$\bar{\mathbf{u}}_{i,g} = (\bar{u}_{0,i,g}, \bar{u}_{1,i,g}, \dots, \bar{u}_{D-1,i,g}), \text{ and}$$

$$\bar{u}_{j,i,g} = \begin{cases} u_{j,i,g} + \sigma(para_{\max}^j - u_{i,i,g}) & \text{if } \sigma > 0 \\ u_{j,i,g} + \sigma(u_{i,i,g} - para_{\min}^j) & \text{if } \sigma \leq 0 \end{cases} \quad (23)$$

$$\sigma = \psi_a(\varphi) = \frac{1}{\sqrt{a}} \psi\left(\frac{\varphi}{a}\right) \quad (24)$$

where the same Morlet wavelet in (15) is used as the mother wavelet and the value of  $a$  is governed by (22). Similar to  $F$  of (20), a larger value of  $|\sigma|$  at the early stage of evolution gives a larger searching space for the solution; when  $|\sigma|$  is small at the later stage of evolution, the algorithm gives a smaller searching space for fine-tuning.



After the operations of the double wavelet mutation, the population is updated by comparing each trial vector  $\bar{\mathbf{u}}_{i,g}$  to the corresponding target vector  $\mathbf{x}_{i,g}$  using the method of standard DE as given by (12). A new population is generated and the same evolution process is repeated. Such an iterative process will be terminated when a defined number of iterations has been reached.

#### 2.2.3.4. DWM-DE for Multi-Objective Optimization

In practice, we often have more than one objective to be optimized in a single problem. A multi-objective optimization problem can be formulated as follows. Optimize  $\{f_1(x), f_2(x), \dots, f_m(x)\}$ , where  $f_i(x); i=1,2,\dots,m$ , is a single objective function;  $m$  is the number of objective functions of the problem.

Optimizing several objectives simultaneously without considering the importance of each objective in the problem is called the Pareto-optimization. Under the Pareto optimization, if a new solution can improve at least one of the objectives without declining the other objectives, this solution will be considered as a better solution than the original one. If the new solution can improve at least one of the objectives, but declining the other objectives, this solution will be considered as a poor solution and will be discarded. The objective of the Pareto optimization is to find a set of solutions that is not dominated by any one solution. This idea can be formulated into the selection operation of DWM-DE as follows:

$$\mathbf{x}_{i,g+1} = \begin{cases} \bar{\mathbf{u}}_{i,g} & \text{if } \forall i f_i(\bar{\mathbf{u}}_{i,g}) \leq f_i(\mathbf{x}_{i,g}) \\ \mathbf{x}_{i,g} & \text{otherwise.} \end{cases} \quad (25)$$

where  $f_i(\cdot)$  is the  $i$ -th objective function. The pseudo code of the resulting multi-objective DWM-DE is shown in Fig. 5.

### 2.2.3.5. Training and Validation

In the proposed system, DWM-DE is employed to optimize the fuzzy inference system (FIS) by finding out the best parameters for the fuzzy rules and membership functions. The job of the FIS is to detect the hypoglycaemic episodes accurately. To measure the performance of the biomedical classification test, sensitivity and specificity are considered. Sensitivity measures the proportion of actual positives that are correctly identified, and specificity measures the proportion of actual negatives that are correctly identified. The definitions of sensitivity ( $\xi$ ) and specificity ( $\eta$ ) are given as follows:

$$\xi = \frac{N_{TP}}{N_{TP} + N_{FN}} \quad (26)$$

$$\eta = \frac{N_{TN}}{N_{TN} + N_{FP}} \quad (27)$$

where  $N_{TP}$  is the number of true positive, which implies the sick people are correctly diagnosed as sick;  $N_{FN}$  is the number of false negative, which implies the sick people are wrongly diagnosed as healthy;  $N_{FP}$  is the number of false positive which implies the healthy people are wrongly diagnosed as sick; and  $N_{TN}$  is number of true negative which implies the healthy people are correctly diagnosed as healthy [14][15]. The values of sensitivity ( $\xi$ ) and specificity ( $\eta$ ) are within 0 to 1. If the value of sensitivity is equal to 1, it means that the classification process is able to classify all the sick people correctly. If the value of specificity is equal to 1, it means that the classification process is able to classify all the healthy people correctly. Sensitivity

and specificity evaluate the performance of the system, which are important criteria for tuning the FIS.

To perform the detection of hypoglycaemic episodes accurately, the proposed system should be able to obtain high values of sensitivity and specificity. The FIS learns the characteristic of the relationship between the system inputs and output through some known dataset. Traditionally, the dataset is divided into two sets, one is used for training the system and the other is used for testing the trained system. However, on using this approach, a phenomenon called overtraining may easily occur. Overtraining refers to the reduction of the generalization ability that can occur as the system is trained. To overcome this problem and to reduce the risk of overtraining, a validation strategy is proposed in this paper. Hence, a validation set is introduced such that the dataset is divided into three sets. While the training set is used to train the FIS, the validation set is used to validate the training result. Then, the testing set is used to test the trained system performance. The main objective of introducing the validation set is to balance the training effect on both the training set and the validation set. As a result, it can avoid the training result being obtained from over fitting with a single data set only.

To train the system with the known dataset, a fitness function should be used to guide the training process and achieve good performance. A proper choice of the fitness function can help the training process achieve a good result. As the validation process is introduced, the number of fitness functions needed to be optimized is increased under the multi-objective approach. We consider four fitness functions for training the FIS.

$$f_1 = \begin{cases} \xi_{target} & \text{if } \xi_{train} \geq \xi_{target} \\ \xi_{train} & \text{otherwise} \end{cases} \quad (28)$$

$$f_2 = \begin{cases} \eta_{target} & \text{if } \eta_{train} \geq \eta_{target} \\ \eta_{train} & \text{otherwise} \end{cases} \quad (29)$$

$$f_3 = \begin{cases} \xi_{target} & \text{if } \xi_{val} \geq \xi_{target} \\ \xi_{val} & \text{otherwise} \end{cases} \quad (30)$$

$$f_4 = \begin{cases} \eta_{target} & \text{if } \eta_{val} \geq \eta_{target} \\ \eta_{val} & \text{otherwise} \end{cases} \quad (31)$$

where  $\xi_{train}$  is the sensitivity of the training set;  $\xi_{val}$  is the sensitivity of the validation set;  $\xi_{target}$  is the target value of sensitivity;  $\eta_{train}$  is the specificity of the training set;  $\eta_{val}$  is the specificity of the validation set;  $\eta_{target}$  is the target value of specificity. The maximum value of each fitness function is equal to 1. These fitness functions are used to train both the training set and validation set for approaching all the target values as much as possible. In the proposed system, the target values of the sensitivity and specificity ( $\xi_{target}$  and  $\eta_{target}$ ) are not constant. They are dynamically updated by some conditions along the search process of the DWM-DE as given by the pseudo code in Fig. 6. At the beginning of the optimization, some initial values are assigned to the two target values, which are both set at 0.10 in this paper. If the sensitivity of both the training set ( $\xi_{train}$ ) and validation set ( $\xi_{valid}$ ) meet the target value of sensitivity ( $\xi_{target}$ ),  $\xi_{target}$  will be increased by 0.01. The same method is also applied to the target value of specificity ( $\eta_{target}$ ). The updating of  $\xi_{target}$  and  $\eta_{target}$  continue until the end of the DWM-DE training and validation process. The objective of this operation is to balance the training effect of both the training set and validation set. For instance, if currently  $\xi_{target} = 0.65$ ,  $\xi_{train} = 0.68$  and  $\xi_{val} = 0.88$ ; according to the training and validation process both  $f_1$  and  $f_3$  are kept at 0.65 and  $\xi_{target}$  for the next iteration is

changed to 0.66. We can see that although  $\xi_{val}$  is much larger than  $\xi_{train}$ , the DWM-DE takes it as the same. As a result, the training process will not bias on the training set or the validation set. The overtraining problem can be alleviated. A more reliable FIS could then be developed.

### 3. Experiment Results and Discussions

Fifteen children with T1DM ( $14.6 \pm 1.5$  years) volunteered to join a 10-hour overnight hypoglycaemia study at the Princess Hospital for Children in Perth, Western Australia. Each patient was monitored overnight for the natural occurrence of nocturnal hypoglycaemia. Data were collected with approval from Woman's and Children's Health Service, Department of Health, Government of Western Australia, and with informed consent.

In this study, we measure the required physiological parameters, while the actual blood glucose levels (BGL) are collected as reference using Yellow Springs Instrument (YSI) blood glucose analyzers. The parameters used for the detection of hypoglycaemia are the HR, QT<sub>c</sub>,  $\Delta$ HR and  $\Delta$ QT<sub>c</sub>. The actual blood glucose profiles for 15 T1DM children are shown in Fig. 10. In this figure, all the data are taken at the same starting time and the duration of data taking for each patient is around 400–480 minutes. The responses from 15 T1DM children show significant blood glucose changes during the hypoglycaemia phase against the non-hypoglycaemia phase. Normalization is used to reduce the patient-to-patient variability and to enable group comparison. It is done by dividing the patient's heart rate and corrected QT interval by his/her corresponding values at time zero.

In this paper, the meaning of hypoglycaemia is suggested as  $BGL < 3.3 \text{ mmol/l}$ . The detection of hypoglycaemia episodes using these input variables is based on a fuzzy inference system trained by the obtained clinical dataset. In effect, it estimates the presence of hypoglycaemia at a sampling period of  $k_s$ , based on the currently sampled data and the data sampled one period before. The sampling period is 5-10 minutes and approximately 35-40 data points from each patient were used. The whole data set is divided into a training set, a validation set and a testing set; each with 5 patients randomly selected. With these 15 patients, the numbers of data points for training, validation, and testing are 199, 177, and 193 respectively. The whole data set, which includes both the hypoglycaemia data part and non-hypoglycaemia data part, is used to measure the classification performance in terms of sensitivity and specificity.

For comparison and analysis purposes, 8 different approaches [10][13][19] are used to tackle the TIDM problem. They are:

- i) The proposed fuzzy inference system with 4 inputs (HR,  $QT_c$ ,  $\Delta HR$  and  $\Delta QT_c$ ) tuned by DWM-DE (FIS-DWM-DE-4) with the proposed training and validation.
- ii) An evolved fuzzy inference system with 2 inputs (HR and  $QT_c$ ) tuned by DWM-DE (FIS-DWM-DE-2) with the proposed training and validation.
- iii) A fuzzy inference system with 4 inputs (HR,  $QT_c$ ,  $\Delta HR$  and  $\Delta QT_c$ ) without validation (FIS-4- w/o-v).
- iv) A fuzzy inference system with 2 inputs (HR and  $QT_c$ ) without validation (FIS-2- w/o-v).

- v) A linear multiple regression with 4 inputs (HR, QT<sub>c</sub>, ΔHR and ΔQT<sub>c</sub>) (LR-4)
- vi) A linear multiple regression with 2 inputs (HR and QT<sub>c</sub>) (LR-2)
- vii) An evolved multiple regressions with 2 inputs (EMR2).
- viii) A feed-forward neural network (FFNN).

In i) and ii), DWM-DE is used to optimize the fuzzy rules and membership functions of the FIS. The settings of the DWM-DE parameters are given as follows.

- Shape parameter of the wavelet mutation ( $\zeta_{wm}$ ): 1
- Parameter  $\lambda$  for the monotonic increasing function: 10000.
- Initial population: It is generated uniformly at random.
- Crossover probability constant:  $C_r = 0.5$
- Maximum number of iteration: 5000
- Population size: 100

For the 2-input FIS, three different numbers of membership functions ( $m_f$ ) have been used for testing:  $m_f = 3, 5$  and  $8$ ; and the numbers of rules are  $9, 25$  and  $64$  respectively. For the 4-input case, only  $m_f = 3$  and  $5$  have been used; and the numbers of rules are  $81$  and  $625$  respectively.

The averaged results out of 50 trials from the proposed system under different numbers of inputs and membership functions are shown in Table 1. The sensitivity and specificity of the training dataset, validation dataset and testing dataset are reported in this table. The  $t_d$  in Table 1 shows the number of parameters for tuning in different cases. As the numbers of inputs and membership functions increase, the numbers of parameters for tuning increase rapidly. It is found that the case of 4 inputs and 5 membership functions offers the best testing result. The best results out of the

50 trials are given in Table 2. We can see that the FIS model can offer more than 75% in sensitivity and 55% in specificity, which are higher than the common clinical classification requirements (sensitivity > 70% and specificity > 50%). As the number of inputs increase, more information could be offered to the FIS to perform better classification. Moreover, the classification performance of the FIS increases when the number of membership function increases thanks to the greater degree of freedom of the FIS. Despite the above advantages, the numbers of parameters for tuning will also increase rapidly. It makes the training process more computationally demanding. The DWM-DE is employed to handle the training process in order to improve the solution quality, solution stability and the rate of convergence.

## 4. Conclusion

In this paper, a fuzzy inference system (FIS) has been developed to recognize the hypoglycaemic episodes. The FIS models the relationship between the heart rate, corrected QT interval of the electrocardiogram (ECG) signal and the hypoglycaemic episodes to perform the classification. Differential evolution with double wavelet mutation operations (DWM-DE) has been employed to optimize the parameters in the FIS. An experiment using data of 15 children with T1DM (569 data points) is studied. To alleviate the effect of overtraining (over-fitting), a validation approach has been proposed. To embed the validation process in the training operation of the FIS, the multi-objective approach of DWM-DE is employed. The classification result shows that the proposed FIS performs well and is able to meet the common clinical classification requirement. To conclude, the FIS could possibly model the relationship between the episodes of hypoglycaemia and the physiological parameters of heart rate (HR), corrected QT interval of the electrocardiogram (ECG) signal



( $QT_c$ ), change of HR and change of  $QT_c$ . There are two limitations for the proposed system, i) for patients with diabetes, the proposed method only works if they have not any heart problems that affect the heart rate and QT interval; otherwise, the performance might be affected; ii) as stated at the end of Section III, the number of parameters for tuning is large when we opt for good performance. This will increase the computational demand for realizing the proposed algorithm.

As directions of future research and development, different system modelling methods could be investigated, which could be tuned by the proposed DWM-DE to enhance the classification performance for hypoglycaemia detection. Example tools might include Support Vector Machines (SVM), Neural Networks and Type-2 Fuzzy Systems. Moreover, other types of physiological signals could be considered in order to make hypoglycaemia detection more reliable. Recently, the Electroencephalography (EEG) signal is under active investigation.

## ACKNOWLEDGEMENT

The work described in this paper was substantially supported by grants from The Hong Kong Polytechnic University (Project No. RP9L) and Juvenile Diabetes Research Foundation, US.

## REFERENCES

- [1] J. F. Yale, Nocturnal hypoglycemia in patients with insulin-treated diabetes, *Diabetes Res. Clin. Pract.*, 65 (2004) S41–S46.
- [2] DCCT Research Group, Adverse events and their association with treatment regimens in the diabetes control and complications trial, *Diabetes Care*, 18 (1995) 1415–1427.
- [3] J. C. Pickup, Sensitivity glucose sensing in diabetes, *Lancet*, 355 (2000) 426–427.
- [4] Directnet Study Group, Evaluation of factors affecting CGMS calibration, *Diabetes Technol. Ther.*, 8 (2006) 318–325.
- [5] F. F. R. Maia, L. R. Araujo, Efficacy of continuous glucose monitoring system (CGMS) to detect postprandial hyperglycemia and unrecognized hypoglycemia in type1 diabetic patients, *Diabetes Res. Clin. Pract.*, 75 (2007) 30–34.
- [6] O. Amir, D. Weinstein, S. Ziberman, M. Less, D. Perl-Treves, H. Primack, A. Weinstein, E. Gabis, B. Fikhte, A. Karasik, “Continuous noninvasive glucose monitoring technology based on ‘occlusion spectroscopy,” *J. Diabetes Sci. Technol.*, 1 (4) (2007) 463–469.
- [7] R. L. Weinstein, Accuracy of the freestyle navigator CGMS: Comparison with frequent laboratory measurements, *Diabetes Care*, 30 (2007) 1125–1130.
- [8] A. Caduff, M. S. Talary, M. Mueller, F. Dewarrant, J. Klisic, M. Donath, L. Heinemann, W. A. Stahel, Non-invasive glucose monitoring in patient with type 1 diabetes: A multisensory system combining sensors for dielectric and optical characterization of skin, *Biosens. Bioelectron.*, 24 (2009) 2778–2784, 2009.

- [9] R. Storn, K. Price, Differential evolution—a simple and efficient heuristic for global optimization over continuous spaces, *Journal of Global Optimization*, 11 (1997) 341–359.
- [10] G. A. F. Seber, *Linear Regression Analysis*. New York: Wiley, 2003.
- [11] B. H. Cho, H. Yu, K. W. Kim, T. H. Kim, I. Y. Kim, S. I. Kim, Application of irregular and unbalanced data to predict diabetic nephropathy using visualization and feature selection methods, *Artif. Intel. Med.*, 42 (1) (2008) 37–53.
- [12] A. Chu, H. Ahn, B. Halwan, B. Kalmin, E. L. V. Artifon, A. Barkun, M. G. Lagoudakis, A. Kumar, A decision support system to facilitate management of patients with acute gastrointestinal bleeding, *Artif. Intel. Med.*, 42 (3) (2008) 247–259.
- [13] S. Wang, W. Min, A new detection algorithm (NDA) based on fuzzy cellular neural networks for white blood cell detection, *IEEE Trans. Inf. Technol. Biomed.*, 10 (1) (2006) 5–10.
- [14] D. A. Freedman, *Statistical Models: Theory and Practice*, Cambridge, U.K.: Cambridge Univ. Press, 2005.
- [15] D. G. Altman, J. M. Bland, Statistics notes: Diagnostic tests 1: Sensitivity and specificity, *Br. Med. J.*, 308 (1994) 1552.
- [16] I. Daubechies, *Ten lectures on Wavelets*, Philadelphia, PA: Society for Industrial and Applied Mathematics, 1992.
- [17] U.K. Chakraborty, *Advances in Differential Evolution*, Springer, Heidelberg, 2008.
- [18] H.Y. Fan, J. Lampinen, A trigonometric mutation operation to differential evolution, *J. Global Optim.*, 27 (1) (2003) 105–129.

- [19] S. H. Ling, H. T. Nguyen, Genetic algorithm-based multiple regression with fuzzy inference system for detection of nocturnal hypoglycemic episodes, *IEEE Trans on Information Technology in Biomedicine*, 15 (2) (2011) 308–315.

Table 1. Results of the proposed system (average of 50 trials).

No of Inputs	$m_f$	$t_d$	Testing		Training		Validation	
			$\xi$	$\eta$	$\xi$	$\eta$	$\xi$	$\eta$
2	3	21	70.37%	40.32%	80.65%	40.08%	86.89%	43.68%
	5	45	72.22%	40.15%	80.66%	43.81%	82.43%	42.64%
	8	96	72.07%	40.71%	81.55%	<b>44.02%</b>	79.00%	43.13%
4	3	105	73.56%	44.98%	82.54%	41.94%	86.30%	45.45%
	5	665	<b>74.92%</b>	<b>53.64%</b>	<b>83.45%</b>	40.66%	<b>90.07%</b>	<b>50.45%</b>

Table 2. Best testing results for hypoglycemic detection from different approaches with sensitivity of around 75%.

Method	Sensitivity	Specificity
FIS-DWM-DE-4	<b>75.92%</b>	<b>55.14%</b>
FIS-DWM-DE-2	74.92%	47.12%
FIS-4-w/o-v	75.00%	51.64%
FIS-2- w/o-v	73.21%	52.58%
LR-4	51.78%	51.64%
LR-2	50.00%	51.17%
FFNN-2	64.26%	52.50%
MR-2	62.31%	53.10%

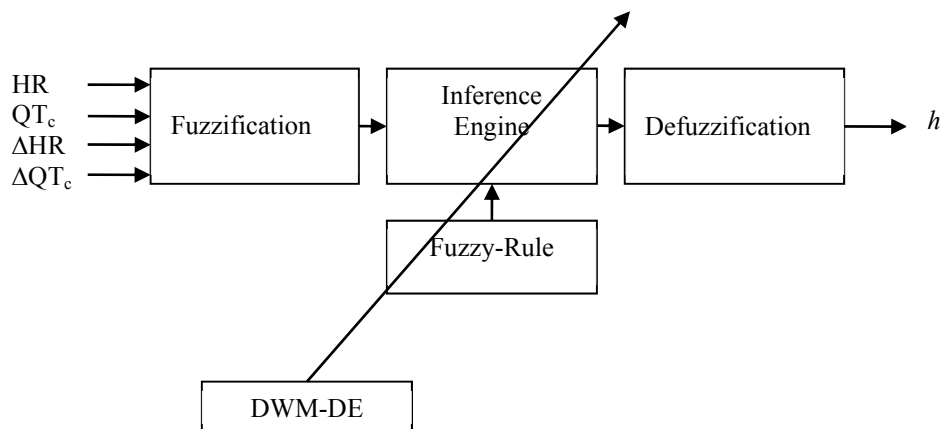


Fig. 1. Fuzzy Inference System (FIS).

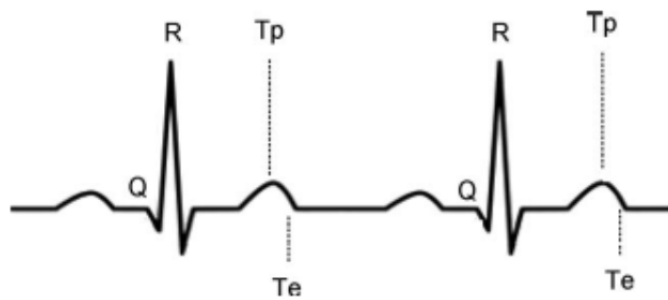


Fig. 2. ECG Signal.

```

begin
  Initialize the population
  while (not termination condition) do
    begin
      Mutation operation by equation (10)
      Crossover operation by equation (11)
      Evaluation of the fitness function
      Select the best vector by equation (12)
    end
  end
end
  
```

Fig. 3. Pseudo code for SDE.

```

begin
Initialize the population
while (not termination condition) do
begin
Mutation operation by equation (10)
Update the new value of  $F$  by equation (20)
Crossover operation by equation (11)
Modifying the trial population vectors based on equation (23)
Evaluation of the fitness function
Select the best vector by equation (12)
end
end

```

Fig. 4. Pseudo code for DWM-DE.

```

begin
Initialize the population
while (not termination condition) do
begin
Mutation operation by equation (10)
Update the new value of  $F$  by equation (20)
Crossover operation by equation (11)
Modifying the trail population vectors based on equation (23)
Evaluation of the fitness functions
Select the best vector by equation (25)
end
end

```

Fig. 5. Pseudo code for Multi-Objective DWM-DE.

```

begin
Initialize the target value
 $\zeta_{target}=0.10$ 
 $\eta_{target}=0.10$ 
while (not termination condition) do
begin
Output ( $\zeta_{train}, \zeta_{val}$ ) by Equation (26) and DWM-DE
Output ( $\eta_{train}, \eta_{val}$ ) by Equation (27) and DWM-DE
Output ( $f_1$ ) by Equation (28)
Output ( $f_2$ ) by Equation (29)
Output ( $f_3$ ) by Equation (30)
Output ( $f_4$ ) by Equation (31)
if ( $f_1=\zeta_{target}$  &  $f_3=\zeta_{target}$ )
then  $\zeta_{target}=\zeta_{target}+0.01$ 
if ( $f_2=\eta_{target}$  &  $f_4=\eta_{target}$ )
then  $\eta_{target}=\eta_{target}+0.01$ 
end
end

```

Fig. 6. Pseudo code for the training and validation.

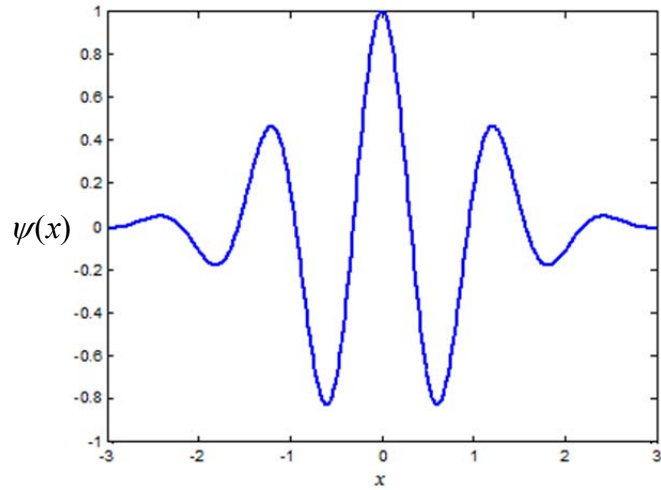


Fig. 7. Morlet wavelet.

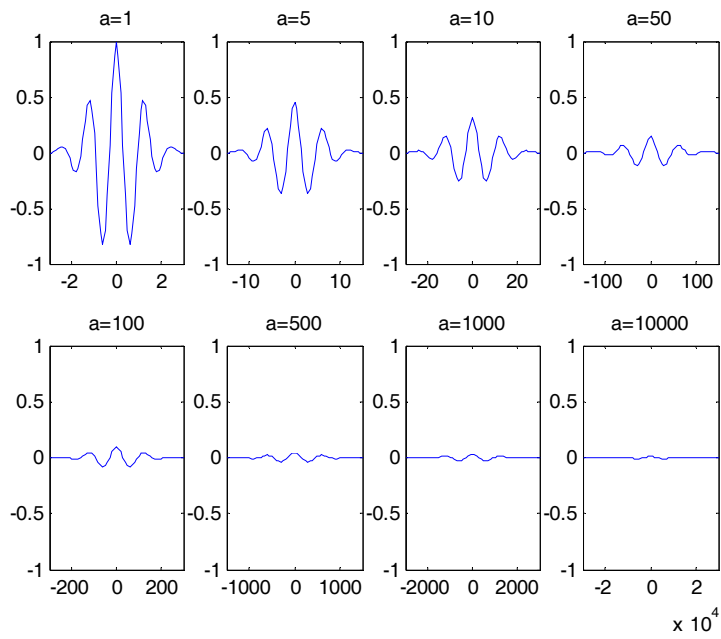


Fig. 8. Morlet wavelet dilated by different values of  $a$  ( $x$ -axis:  $a$ ,  $y$ -axis:  $\psi_a(x)$ .)



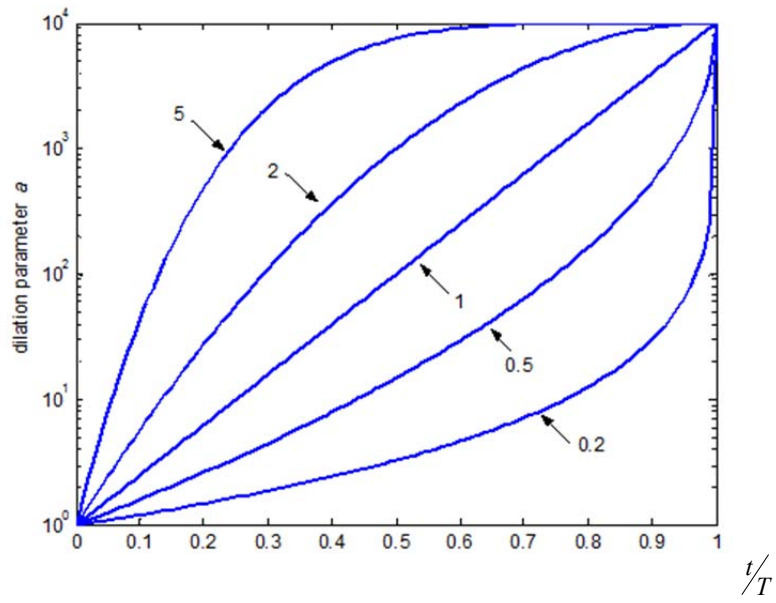


Fig. 9. Effect of the shape parameter  $\zeta_{wm}$  to  $a$  with respect to  $t/T$ .

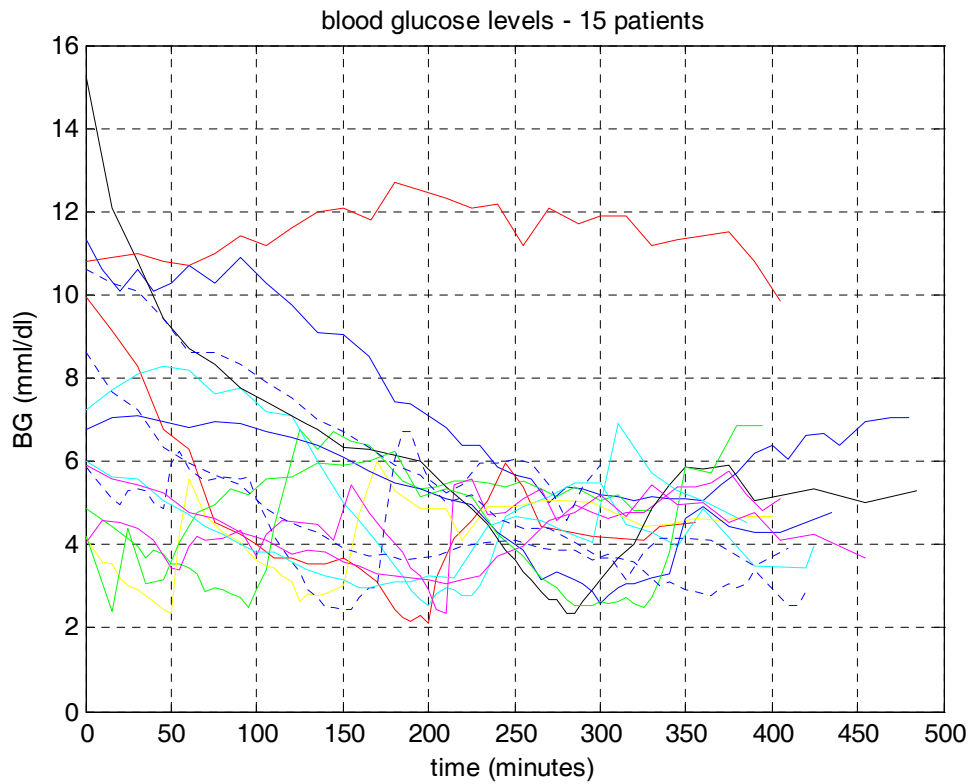


Fig. 10. Actual BG-Level profiles in 15 T1DM children.