Fuzzy Inference Model for Type 2 Diabetes Management: a Tool for Regimen Alterations

Nonso Nnamoko¹*, Farath Arshad¹, David England¹, Jiten Vora², James Norman³

¹School of Computing and Mathematical Science, Liverpool John Moores University, Liverpool, United Kingdom
²Diabetes and Endocrinology department, Royal Liverpool and Broadgreen University Hospitals, Liverpool, United Kingdom
³Information Management and Technology department, Royal Liverpool and Broadgreen University Hospitals, Liverpool, United Kingdom

*Corresponding author: N.A.Nnamoko@2011.ljmu.ac.uk

Received May 07, 2015; Revised July 08, 2015; Accepted July 16, 2015

Abstract This paper aims to demonstrate the utility of fuzzy set theory in the design process of a diabetes management system that enables patients to make short term alterations (particularly lifestyle) to their overall regimen as required. The model is a Mamdani Fuzzy Inference System (FIS) configured through domain specific information from experts and recognised diabetes management algorithms. The FIS takes a multi-input multi-output (MIMO) design approach with seven inputs variables (age, gender, weight, height, blood glucose (BG), exercise and diet) and three outputs (glycatedhaemoglobin (A1c), exercise and diet level assessments). Goodness of fit test was conducted based on Mean Square Error (MSE), Normalised Mean Square Error (NMSE) and Normalised Root Mean Square Error (NRMSE) between observed/advised and predicted output values. Overall MSE of 0.0899 shows good fit. For each of the output pairs (A1c, exercise and diet), NRMSE (0.7387, 0.7881 and 0.3716) and NMSE (0.9317, 0.9551 and 0.6051) shows good fit for A1c and exercise, but poor fit for diet. Intelligent models of this sort can help simplify management information for diabetes patients, reduce routine workload for clinicians and allow them to focus more on critical issues. Fully developed, this system can be used to build a database of diabetes management cases that includes daily life event information, ultimately leading to automated care for patients through technology.

Keywords: fuzzy logic, diabetes management, fuzzy inference system, rule based reasoning, case based reasoning


1. Introduction

Type 2 diabetes management requires a strict diet and exercise regimen agreed upon between patients and clinicians in order to stabilise blood glucose (BG) level and maintain good health [1]. Conventional diabetes care delivery adopts a long-term feedback system, involving structured educational programmes and care plan adjustments during routine checks (usually 3 – 6 months intervals). However, the monitored indicators are dynamic in nature, requiring short-term alterations to compensate for any change in a patient’s scheduled regimen. It is difficult for a patient to ascertain when a change is required, the nature of the change, time to implement any change, the required magnitude or even the resultant effect of any change applied. Even more difficult is developing a computational model that deals with these issues, due to the numerous metabolic reactions that occur within the human body [2]. For instance, changes in glucose concentration are dependent upon heart rate (HR) [3,4], insulin sensitivity [5,6], glucose production by the liver [7], and a myriad of hormonal interactions [8]. Artificial Intelligence (AI) holds great promise in resolving this issue. The model described in this paper, based on Mamdani Fuzzy Inference System (FIS) was configured using expert knowledge and management algorithms calibrated through large datasets of patient specific information. The model is intended to provide short-term (daily) interpretation of monitoring data, in form of personalised decision support for achieving personal health goals. That is, turning endless daily data into useful insight so immediate changes can be made to compensate for any regimen disturbance. Seven input variables were used including patient demographic information (age and gender), patient specific information (weight and height) and daily regimen event types (i.e., calorie intake, HR and BG. To reduce the number of rules required to structure the model appropriately, the weight and height variables were interlaced with the other five variables using equations from related research studies to enable personalised outputs.

2. Context

Diabetes is a hereditary or developmental condition caused by the malfunctioning of the pancreas, which
secretes the hormone insulin, resulting in elevated glucose concentration in the blood. In some cases, the body cells fail to respond to the normal action of insulin (i.e., insulin resistance). There are two main types of diabetes – Type 1 and Type 2. Type 1 is the least common, developed when the body cannot produce any insulin and falls outside the scope of our study. Type 2 however, develops when the body can still produce some insulin, but not enough is made available. For many diagnosed cases of type 2, the first approach to treatment is through BG monitoring several times a day, and lifestyle adjustments to improve readings [9]. In the UK, this treatment approach is delivered through structured educational programmes, care planning and routine checks (3 – 6 months intervals) to identify and apply necessary changes to a course of treatment regimes. A common educational programme applicable to type 2 cases is DESMOND; an acronym for diabetes education and self-management for on-going and newly diagnosed [10].

As noted earlier, management may be compromised due to insufficient data and in some cases, patients’ inability to interpret available data. However, several research attempts using AI techniques have been reported that may remove some of these burdens from patients. One notable example is the automated insulin dosage advisor (AIDA), a mathematical model intended for type 1 (insulin dependent) diabetic patients to simulate the effects of changes in insulin and diet, on BG profile [11,12]. However, the authors made clear the limitations of this model and declared it insufficiently accurate for patient use in BG – insulin regulation. Despite this limitation, the authors believe there is still value in its capability as an educational tool for diabetics, carers or even researchers. For instance, the artificial neural network (ANN) based model for BG prediction by Robertson et al. [6] was trained with simulated data from AIDA. Their main contribution is to quickly detect hyperglycaemic or hypoglycaemic episodes from BG predictions, using insulin dosage and diet. Dazzi et al. [13] presented a model also aimed at type 1 diabetes management with similar contribution using neuro-fuzzy method. Their aim was to adjust insulin dosage based on desired BG levels.

Whereas type 1 diabetes has received a lot of interesting research in terms of electronic management solutions, type 2 (non – insulin) diabetes is still under researched; perhaps due to the numerous factors to consider in its management. In type 2 diabetes management, emphases are more on BG interaction with lifestyle (exercise and diet) data, and not the usual insulin dosage data as in type 1 cases. Majority of published solutions in this area are either medically led with no computational element [14-18], or presented as a black box [19-23]; thus internal computational elements or approach were hidden and validation were mostly descriptive based on user experience [20,21]. In this paper, we present a detailed explanation and validation of a model, intended for early type 2 (non - insulin) diabetes management. The model uses a Mamdani Fuzzy Inference System (FIS) configured through domain specific information from experts [24,25,26] and recognised diabetes management algorithms [27,28,29]. It is important to note that this study is not intended to review or compare existing AI approaches that may be applicable in this domain. Rather, the authors aim to describe and validate explicitly, the utility of FIS in dealing with the incomplete and disjoint datasets involved in type 2 diabetes management. Details of the internal configuration are presented in the next section.

### 3. Method

We used FIS because of its capability in dealing with data imprecision and quantification that are common in diabetes management. For instance, BG values from glucose monitors are not particularly precise and patients might find it difficult to quantify and/or associate the effects of the BG values with the desired health outcome, given that other life events (e.g., exercise and diet) have to be taken into account. FIS formalises these values by grouping them into fuzzy sets using Membership Functions (MF); and forming relationships with the fuzzy sets to produce outputs. Basically, using fuzzy clusters of similarity (fuzzy sets) we can hide unwanted values, ultimately leading to systems where the grouped values can be used to focus analysis on aspects of interest to the user. For instance, to determine the effect of a person’s BG value towards the desired output, MFs would locate where the value lies within the fuzzy sets and the relationships would locate the assigned output space that defines the output achieved. The fuzzy sets exhibit two different shapes for membership functions (triangular & trapezoidal) as shown in Table 1. The relationships are sets of fuzzy rules (n = 165) used for inference and are stored within the FIS knowledge base (see Figure 1). Like the model described in this paper, some FIS are based on rules obtained by eliciting information from experts [30], although data driven approaches exist where rules are extracted automatically from empirical data based on trends [31,32]. The rules are based on expert knowledge in diabetes management [24,25,26]. The parameters used for fuzzy set classification are based on official research studies about the characteristics of the input variables in diabetes management [27,28,29].

FIS usually take two forms, namely: multiple-input multiple-output (MIMO), where the system returns several outputs based on the inputs; and multiple-input single-output (MISO), where only one output is returned from multiple inputs [33]. The model presented in this paper is based on MIMO with seven inputs and three outputs. The inputs include five primary variables (age, gender, HR, BG and calorie intake) and two auxiliary variables (weight and height). Table 2 shows the input variables with associated fuzzy set classifications. The determined output variables namely; exercise level, calorie level and glycated haemoglobin (commonly known as HBA1C or A1C) are shown in Table 3, each with a set of linguistic values. The auxiliary variables are nested into the primary ones with a set of rules within the FIS knowledge base.

We determined exercise level based on percentage of maximum HR, calculated using equation (1) by Keytel et al. [24]; and classified in accordance with the compendium of physical activity measures as reported by Ainsworth et al. [25] and Mayo Clinic recommendations [29].

\[
\text{Maximumheartrate} = 208 - (0.7 \times \text{Age})
\]
We determined A1c based on daily average BG measurements; calculated and classified in accordance with the A1C-Derived Average Glucose (ADAG) linear equation (2) by Nathan et al. [27]. This representation will help patients more directly see the difference and/or relationship between their daily BG measurements and their overall glucose management performance in terms of A1c.

\[ 28.7 \times \text{A1C} + 46.7 = eAG \]  

(2)

We determined calorie level based on current UK calorie consumption guidelines [34], i.e., 2000 calories for women and 2550 for men. However, these values can vary depending on individual factors such as age, gender, size (i.e., weight and height) and physical activity level (PAL). For personalisation, we used estimated average requirement (EAR) to determine calorie level; calculated using energy equations (3) and (4). The output will help patients more directly see the difference and/or relationship between their daily calorie consumption and recommended limits.

\[ \text{EAR} = \text{RMR} \times \text{PAL} \]  

(3)

PAL values were adopted from the most recent FAO report on human energy requirements [26]. RMR (Resting Metabolic Rate) values representing the energy required for body metabolism while at rest were calculated from Mifflin-St Jeor [28] predictive energy equation (4) at weights equivalent to a body mass index (BMI) of 22.5 kg/m² and current mean heights for age derived from Health Survey for England 2012 [35].

\[ m\_\text{RMR} = 66.47 + (13.75 \times \text{wt}) + (5 \times \text{ht}) + (6.76 \times \text{age}) \]

\[ f\_\text{RMR} = 655.1 + (9.56 \times \text{wt}) + (1.85 \times \text{ht}) + (4.68 \times \text{age}) \]  

(4)

Although these values are for average adults (not diabetes patients), our decision to use them follows the concept that healthy nutrition recommendations for the general public are also appropriate for persons with type 2 diabetes [36]. We preferred the Mifflin-St Jeor equation over others because it gives the most reliable result [37].

Table 1. Triangular and Trapezoidal Fuzzy shapes with associated MFs

<table>
<thead>
<tr>
<th>Shape</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Triangular</td>
<td>[ \mu(x) = \begin{cases} 0, &amp; x \leq a \ \frac{x-a}{b-a}, &amp; a \leq x \leq b \ \frac{c-x}{c-b}, &amp; b \leq x \leq c \ 0, &amp; c \leq x \end{cases} ]</td>
</tr>
<tr>
<td>Trapezoidal</td>
<td>[ \mu(x) = \begin{cases} 0, &amp; x \leq a \ \frac{x-a}{b-a}, &amp; a \leq x \leq b \ 1, &amp; b \leq x \leq c \ \frac{d-x}{d-c}, &amp; c \leq x \leq d \ 0, &amp; d \leq x \end{cases} ]</td>
</tr>
</tbody>
</table>

Figure 1. The Fuzzy Inference System (FIS)
Table 2. Input variables and associated fuzzy set classifications

<table>
<thead>
<tr>
<th>Input Variables</th>
<th>Fuzzy set classifications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>0 = “Male” and 1 = “Female”</td>
</tr>
<tr>
<td>Age</td>
<td>≤ 35 = “Young”; 30 – 50 = “Mid-age”; ≥ 50 = “Old”</td>
</tr>
<tr>
<td>HR</td>
<td>≤ 55% = “Light”; 50 – 75% = “Moderate”; ≥ 70% = “Vigorous”</td>
</tr>
<tr>
<td>BG</td>
<td>≤ 7.2 = “7”; 6.8 – 8.4 = “7.8”; 8.2 – 9.0 = “8.6”; 9 – 9.9 = “9.4”; 9.7 – 10.7 = “10.1”; 10.5 – 11.6 = “10.9”; 11.4 – 12.4 = “11.8”; 12.2 – 13.2 = “12.6”; ≥ 13 = “13.4”</td>
</tr>
</tbody>
</table>

Table 3. Output variables and associated linguistic values

<table>
<thead>
<tr>
<th>Output Variable</th>
<th>Linguistic Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1c</td>
<td>A1c values between 4 – 12%</td>
</tr>
<tr>
<td>Exercise level</td>
<td>&lt; Limit; Within guide limit; &gt; Limit</td>
</tr>
<tr>
<td>Calorie level</td>
<td>&lt; Limit; Within guide limit; &gt; Limit</td>
</tr>
</tbody>
</table>

4. Theoretical Usage Scenario

In practice, users would be presented with a web based user interface where the required input data can be entered and outputs displayed as shown in Figure 2. Gender rarely changes and age runs with calendar, so single entry is required for both input variables. However, periodic entries would be required for the remaining five input variables; six months for weight and height, and daily (preferably before bed time) for calorie intake, average HR and BG.

Outputs are displayed for each entry period in form of advice to indicate any issue(s) and/or required alteration(s) to regimen. In addition, the FIS updates the database with every entry and corresponding outputs for each user. Such database would be useful in future improvements to the model and facilitate automated decision support as proposed in our previous research work [38].

5. Evaluation and Limitations

Evaluation was limited due to resource constraints. Small scale evaluation of the model capability was conducted using five weeks monitoring data of 35 subjects with type 2 diabetes, including continuous BG measurements (CBGM) data from Abbott Freestyle Navigator II [39]. The dataset includes 48 features with blind assessments of achievements. Only the relevant features for this evaluation (i.e., age, gender, weight, height, HR, Diet, CBGM and A1c) were extracted with advised achievement levels for diet and exercise. For BG input variable we used average daily CBGM. Weight, height, HR and diet values were only monitored once within the five week period, so we only used one day extract per subject for this evaluation. We used the supplied A1c values along with diet and exercise assessments as outputs. Details of assessment method were not supplied with the data.

To establish how well the model performed, we conducted goodness of fit tests [40] between predicted values x and observed/advised values y, using the following cost functions:

1. Mean square error (MSE)
2. Normalised root mean square error (NRMSE)
3. Normalised mean square error (NMSE)

MSE, NRMSE and NMSE costs vary between $-\infty$ (bad fit) to 1 (good fit). Figure 3 shows plots of the measured data pairs. If the cost function is equal to zero, then x is no better than a straight line at matching y.

The MSE measured as the average sum of the squares of the difference between the estimated and actual values is given by (5), where x is a vector of n predictions, and y is the vector of the true values.

Figure 2. Context Diagram of the model usage

Figure 3. Advised Vs Predicted Output
\[ \text{MSE}_{(x,y)} = \frac{1}{n} \sum_{i=1}^{n} (x_i - y_i)^2 \]  

(5)

Results show overall MSE of 0.0899 which indicates minimal variance between the output data pairs, thus a good fit. Table 4 shows the NMSE and NRMSE results. The NMSE, derived from MSE is given by (6).

\[ \text{NMSE}_{(x,y)} = \frac{\text{MSE}_{(x,y)}}{\text{MSE}_{(x,0)}} \]  

(6)

The NRMSE incorporates some of the embedded variability of MSE by using the square root of MSE (i.e., RMSE) as shown in (7) and (8).

\[ \text{RMSE}_{(x,y)} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - x_i)^2} \]  

(7)

\[ \text{NRMSE}_{(x,y)} = \frac{\text{RMSE}_{(x,y)}}{y_{\text{max}} - y_{\text{min}}} \]  

(8)

<table>
<thead>
<tr>
<th>Output Variables</th>
<th>NMSE</th>
<th>NRMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1c</td>
<td>0.7387</td>
<td>0.9317</td>
</tr>
<tr>
<td>Exercise Level</td>
<td>0.7881</td>
<td>0.9551</td>
</tr>
<tr>
<td>Diet Level</td>
<td>0.3716</td>
<td>0.6051</td>
</tr>
</tbody>
</table>

Table 4. Goodness of Fit evaluation

From the table, results show good fit for A1c and Exercise level predictions. However, results for Diet level were very poor. This result could be due to a number of reasons e.g., test data veracity, assessment criteria used by Abbott etc.

6. Conclusion and Future Work

Intelligent models of the type discussed can help simplify management information for diabetes patients. This could reduce routine workload for clinicians and allow them to focus more on critical events. During routine check-up, people with diabetes would normally receive around 15 – 20 minutes of interaction with clinicians who must quickly evaluate the patient’s health status and offer therapeutic advice. This time is certainly not sufficient, considering the volume of management data that needs to be processed. Also, clinicians would normally rely on self-reported information of preceding lifestyle behaviour from patients. Whereas BG readings are automatically recorded during daily monitoring and so readily available, lifestyle data are rarely recorded daily by patients. The model described in this paper would enable more efficient data collection from patients, and logical decision from the care team. A possible future for the model is in building a database of diabetes management data that includes daily life event information. Such level of data tracking is known to yield better results [41] and would prove invaluable in developing a case repository for automated case based decision support systems as proposed in our previous research work [38]. Of course such systems would (in some cases) lack clinical intuition especially when presented with new or complex cases. However, the ability to learn and adapt gives these systems greater potential for dealing with patients faster than the conventional support method, ultimately leading to more efficient healthcare support from the care team.

The model described in this paper is a theoretical demonstrator of a planned system. Developing and testing the model in a real user environment will be the next step. In the future, patients will be able to upload their data online and monitor outcomes. Data acquired from this would be utilised in the second phase of the overall project (see Figure 1 in [38]) to provide support for clinicians.

Acknowledgement

We would like to thank the Royal Liverpool & Broadgreen University NHS Hospitals Trust (RLBUHT) and the Faculty of Technology & Environment at Liverpool John Moores University for providing full funding for this PhD research work. We also want to thank Abbott Diabetes Care, UK for providing data for this work. Professor JitenVora, consultant in Diabetes and Endocrinology at RLBUHT provides clinical supervision for this work.

References


