Fuzzy Inference System for Risk Evaluation in Gestational Diabetes Mellitus

Carlos Salort Sánchez*, Suzanne Smyth†, Elizabeth Tully‡, Joanna Griffin§,
Luke Heaphy¶, Niamh Redmond‖, Fionnuala M Breathnach**, Jan Baumbach††, and Cristian Axenie‡‡
IT Software Infrastructure Lab, *Huawei German Research Center*, Munich, Germany
Email: *carlos.salort@huawei.com, ‡‡cristian.axenie@huawei.com
† ** *Royal College of Surgeons in Ireland*, *Rotunda Hospital*, Dublin, Ireland
‡ ¶ ‖ *RCSI Department of Obstetrics and Gynaecology*, *Royal College of Surgeons in Ireland*, Dublin, Ireland
§ *Department of Research*, *Rotunda Hospital*, Dublin, Ireland
* †† *TUM School of Life Sciences Weihenstephan*, *Technical University of Munich (TUM)*, Freising, Germany

***Abstract*—Remote monitoring health data analysis holds the potential to reduce pregnancy complications, improve patients' quality of life, enhance the efficiency of healthcare delivery and reduce healthcare costs. In this paper, we present a method based on fuzzy inference systems to monitor pregnancies complicated by gestational diabetes mellitus (GDM). The system is simple, fast, flexible and exploits domain expertise in assessing risk levels according to capillary glucose levels from women with GDM. We show that this approach generates an interpretable input, which is valuable in medical applications. To prove the capabilities of the system, we present prediction results from 50 real-world patients and show that the system obtains relevant glycaemic-control data comparable to current monitoring methods that rely on periodic face-to-face physician review. Our systems achieves 95% accuracy. Moreover, we show that the difference in predictions account for a more personalized treatment.**

***Index Terms*—E-Health, Fuzzy Inference System, Gestational Diabetes Mellitus**

I. Introduction

There are approximately 60 million people living with diabetes in Europe, accounting for approximately 10.3% of men and 9.6% of women over the age of 25. Accordingly, rates of gestational diabetes mellitus (GDM) are also increasing rapidly due to unhealthy lifestyle, increased obesity prevalence [1] and a lower thresholds for diagnosis in recent years. As a result, many more women are faced with diabetes during pregnancy which puts both mother and child at risk of developing type II diabetes later in life [2]. The International Diabetes Federation (IDF) has estimated that in Europe 89 billion euro were spent on treating and managing diabetes and its related complications in 2011 (work days lost not included).

A more efficient glucose monitoring system and early detection of GDM will positively impact a patient's quality of life and holds the potential to confer significant healthcare cost savings and improved efficiency in clinic processes. At the Rotunda Hospital (https://rotunda.ie/), a tertiary referral major perinatology centre in Dublin, approximately 9000 babies are delivered every year. Despite this high-throughput environment, many services continue to rely on outdated, paper-based, information systems. Currently at the Rotunda, pregnancies complicated by diabetes are managed alongside routine antenatal care in an outpatient department with limited staff and resources. The diabetes care team sees approximately 25 patients in hospital every day for visits that require the patient to remain in hospital for up to 3 hours. In addition to running the clinics, midwives are responsible for receiving approximately 100 calls/week from patients to relay their blood sugar levels over the phone. The midwives manually transcribe this information and then pass it on to a consultant. This process is overly manual, time-consuming and vulnerable to human error. Furthermore, evidence has shown that as many as 22% of women with gestational diabetes falsify glucose values. This is a recognized challenge faced by the health care team and can prevent appropriate management which in turn puts these women and their babies at risk. The application of Big Data Analytics solutions to assist with the secure transfer, storage and analysis of health information holds the potential to greatly improve current approaches. BigMedilytics (Big Data for Medical Analytics) is the largest EU-funded initiative to transform the regions healthcare sector by using state-of-the-art big data technologies to achieve breakthrough productivity in the sector by reducing cost, improving patient outcomes and delivering better access to healthcare facilities simultaneously. One of the pilots within this project (https://www.bigmedilytics.eu/pilot/diabetes/) aims to develop a complete monitoring system, including a mobile app (connected with glucometers to ease data collection, and presenting analytics results to the patients), and a web portal to present the data to the medical team . This work presents the initial model used for monitoring.

In order to extract the underlying correlations and insights of the processes governing GDM data (i.e. glucose level) data from 50 patients were collected and analysed. With the advent of Big-Data and modern Machine Learning (ML) and Artificial Intelligence (AI) algorithms, such data mining applications are sought after, both by research and clinical labs [3]. Expert systems, as a branch of AI, are used when applying specialized knowledge [4]. Such systems use this knowledge to make

The paper is a result of a Horizon 2020 innovation programme under grant agreement No. 780495, project BigMedilytics (Big Data for Medical Analytics).

2471-7819/19/$31.00 ©2019 IEEE
DOI 10.1109/BIBE.2019.00177

suggestions to users who may not have the full range of expertise available to the system. These systems are capable of manipulating symbolic as well as numerical data and have the ability to interact with the user in something approximating natural language, such as fuzzy logic [5]. Moreover, such systems are used as a core element of Decision Support Systems (DSS), which are typically the clinical end-points [6]. In GDM, DSS form a significant component of clinical knowledge management technologies through their capacity to support the clinical process and the use of knowledge, from diagnosis and investigation through treatment and long-term care. Diagnosis needs the integration of different sources of data (i.e. glucose reading from the patient) and the on-line or off-line collaboration of different kinds of specialists (e.g. obstetricians and endocrinologists). DSSs are typically designed to integrate a medical knowledge base, patient data and an inference engine to generate case-specific advice. Current DSS systems exploit expert systems implementations that eliminate the uncertainty and imprecision associated with the diagnosis of gestational diabetes, using fuzzy classification [7]. Despite the flexibility of the proposed system, this approach only used artificial data and considered a large number of features to monitor, which in real-world scenarios might not be available or may be expensive to obtain due to the lack of sensors. Using an adaptive neurofuzzy inference system (ANFIS), the study in [8], combined the regression power of neural networks and flexibility of fuzzy logic in a system that has short training time and average accuracy in prediction. The drawbacks that such a system carries is the lack of explainability and the training time which will be triggered whenever new patient data is available. Moreover, none of the reviewed studies are addressing real-time continuous glucose monitoring in gestational diabetes [9], as such a task requires a lightweight, inference system with flexible use of domain expertise that offers a timely and interpretable output. Another aspect motivating our work looks at the need for making a trade-off between which glycaemic markers are informative to be monitored for fast risk assessment [10]. The proposed approach for GDM monitoring exploits the available real-world data of 50 patients and provides results comparable to the output of a specialized clinician. The output can be easily embedded in the clinical infrastructure or in personal monitoring systems. With a pragmatic approach targeted at delivering a flexible and efficient prediction mechanism, our Fuzzy Inference System obtains comparable output with the physician opinion. The system is employing a simple and explainable structure, fast and adaptive processing (no training), and offers an interpretable output tailored for each patient. We believe that such an approach can contribute to the current initiatives targeting personalized medicine.

II. Materials

In this section we will describe the data and the experimental instrumentation used to validate our model.

A. Data analysis

The data used for validating the model was collected in a study in Rotunda hospital, Dublin. GDM patients are diagnosed approximately in the 28th pregnancy week. Participating patients were asked to collect four daily glucose level measurements until delivery, one in the morning (fasting) and the remaining three readings taken one hour after the three main daily meals (postprandial). For this analysis, fully anonymized data from 50 patients were used, with different levels of adherence to the schedule of measurements, ranging from one every week (i.e. around 10 measurements in the whole period), to several daily measurements (i.e. 700 measurements).

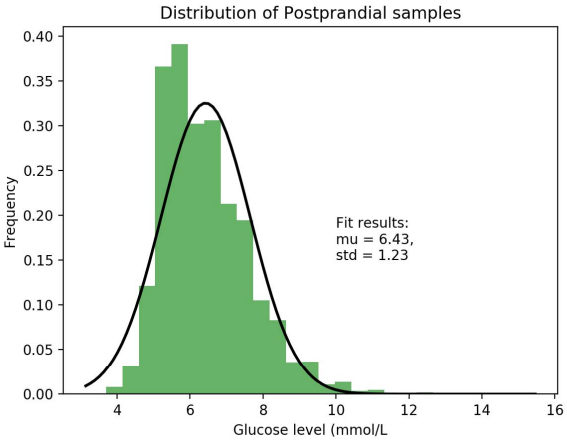

Fig. 1. Distribution of postprandial glucose levels.

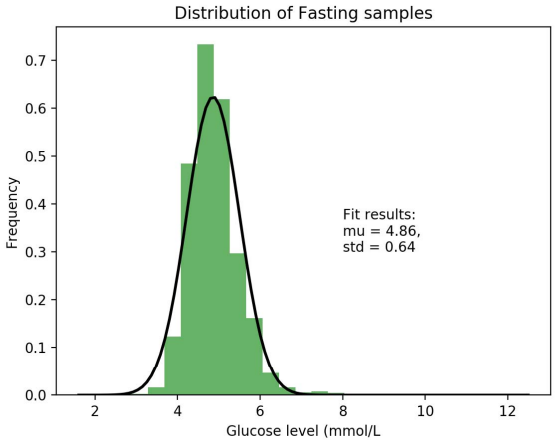

Fig. 2. Distribution of fasting glucose levels.

The data distribution can be observed in figures 1 and 2 for postprandial and fasting measurements respectively. We fit a normal distribution to the values observed, which we will use in section IV to validate the model. Original data

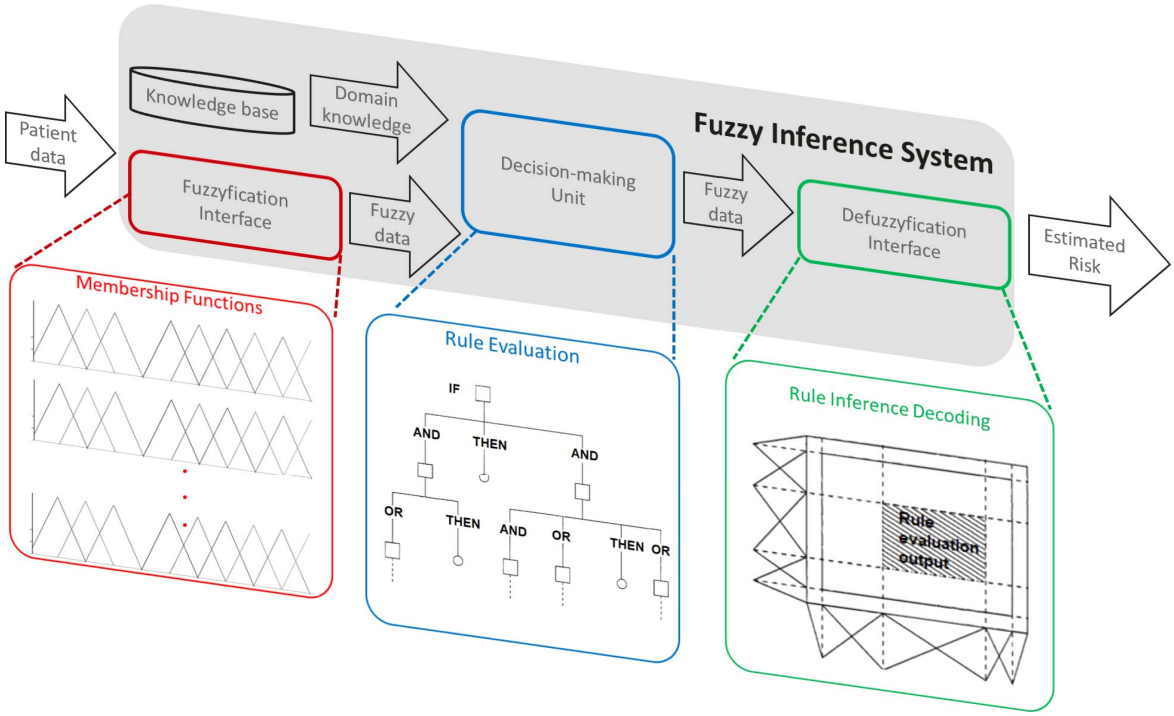

Fig. 3. Generic architecture and processing of the Fuzzy Inference System.

are used exclusively to fit the distributions that will generate the validation samples.

B. Fuzzy Inference Rules Design

For the definition of our fuzzy inference expert system, we used protocols collected from different hospitals in Ireland. In order to monitor GDM, hospitals currently analyze one week of data, and call the patient in when more than 30% of the values are over a certain threshold. The thresholds depend on the hospital, and they are different for postprandial and fasting measurements. These thresholds range around 7 mmol/L for postprandial measurements, and 5 mmol/L for fasting measurements. We will use this values as a base for the fuzzy inference rules definition.

III. Methods

A. Fuzzy Inference Systems

1) Core principle: Fuzzy inference systems are regarded as modeling structures with well-defined functional blocks of input and output interfaces along with a processing module that carries out all the computation at the linguistic level. Generically, a fuzzy inference system consists of four parts: the fuzzification interface, knowledge base, decision-making unit, and defuzzification interface, as shown in Figure 3.

The essential role of the fuzzification is to convert the information coming from the environment (numerical quantities) in an internal format (i.e. through membership functions) acceptable by the knowledge-base and the decision-making unit (i.e. evaluating IF-THEN statements). Symmetrically, the defuzzification provides a conversion procedure to transform information coming from the decision-making module into the form acceptable by the modeling environment (i.e. interpreting the rule evaluation output as a numerical quantity), as shown in Figure 3. The semantic integrity enables humans to assess the meaning of the linguistic terms formalized by a set of rules that should be considered in the design of the input and output interfaces [11].

2) Interpretability: Fuzzy Inference relies on linguistic representation of knowledge that is processed by operating at the semantical level defined through fuzzy logic. Linguistic representation of knowledge is a core feature of fuzzy systems where rule models are acquired from data supporting the overall interpretability of such a system [12]. Yet, more formal and grounded approaches were developed for evaluating interpretability. The interpretability of a fuzzy rule-based model is measured in terms of cointension degree between the explicit semantics, defined by the formal parameter settings of the model, and the implicit semantics conveyed to the reader by the linguistic representation of knowledge [13].

3) Explainability: The objective of explainability is boosting the transparency in the solutions proposed for applications, and therefore the ability to trust the system output. A huge motivation for our approach is rising also from the legal and privacy aspects. The new European General Data Protection Regulation that became active on May 25th 2018, requires to be able to give meaningful information about the decision making (art. 14 of GDPR). This does not imply a ban

on ML approaches or an obligation to explain everything all the time, however, there must be a possibility to make the results re-traceable on demand. Fuzzy Inference System incorporate explainability directly into the structure of the model, hence achieving explainability by design. Moreover, as an interpretable glass-box approach, fuzzy inference has a long tradition providing a good framework for the interaction between human expert knowledge and hidden knowledge in the data [14].

B. Fuzzy Inference Systems for GDM

The proposed Fuzzy Inference System uses a one-week (7 days) rolling window for monitoring patients diagnosed with GDM. At a rate of 4 measurements per day (one fasting, three postprandial), 28 different input variables are considered. The fuzzy system collects the input values for the week, and assigns a risk level between 1 and 100 to the observation. Finally, the system evaluates the rules in order to generate the output based on the inputs configuration. Every input has two membership functions (i.e. LOW and HIGH), and thus defining explicitly all rules would generate 2^{28} different combinations. As this is not feasible from a computational standpoint, we will apply a rule pruning algorithm to reduce the rules set.

1) Input variables: Input variables are separated in two groups with different thresholds values: fasting and postprandial. To create the entire input configuration, we first stack the postprandial values, and then the fasting values. We define

$$in := [post_0, ..., post_{npost-1}, pre_0, ..., pre_{npre-1}] \quad (1)$$

as the input vector for the crisp (real) values for the system.

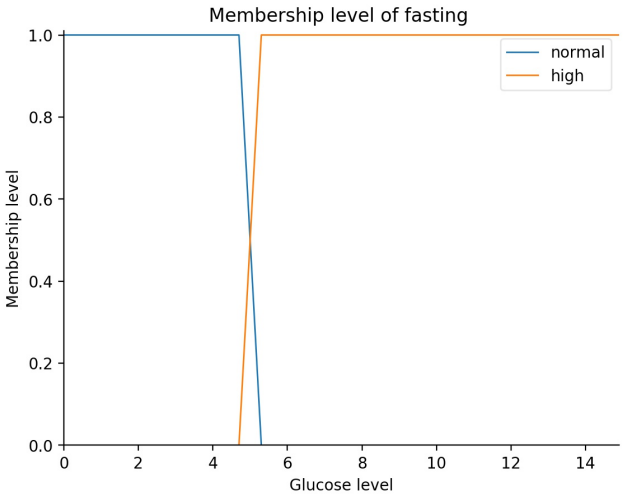

Fig. 4. Membership functions for fasting measurements.

Input values are considered to range in the interval $[0, 15]$. This range includes all glucose level detected in the original study.

In order to define the membership functions for fuzzification, we used the thresholds currently used in GDM monitoring. We took the most used value (i.e. 7 for the postprandial measurements, and 5 for fasting), and gave some variability (i.e. var in the formulas) so that it would include rules from all hospitals. For this paper we use $var = 0.3$.

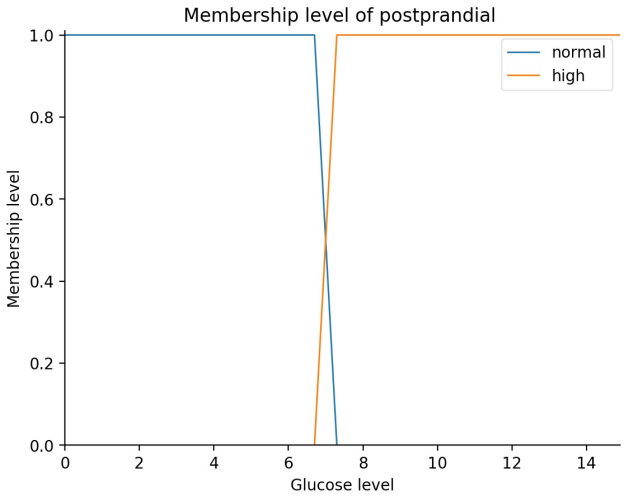

Fig. 5. Membership functions for postprandial measurements.

Membership functions can be found in figures 4 and 5. They use a trapezoidal shape, defined as

$$normal = [0, 0, threshold - var, threshold + var]$$

$$high = [threshold - var, threshold + var, 15, 15]$$

The values represent the left lower and upper, and right upper and lower values of the trapezoid respectively.

2) Output variables: The objective of our system is to assess the risk factor on the integer interval $[1, 100]$. In order to achieve this, we modeled the risk as shown in figure 6.

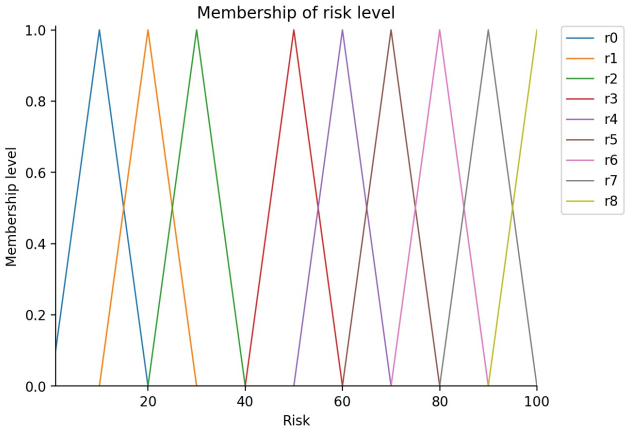

Fig. 6. Membership functions for the output variable (risk).

The membership functions are triangular, and defined as

$$r(i) = [i, i+10, i+20], i = 0, 1, 2$$

$$r(i) = [i+10, i+20, i+30], i = 3, 4, 5, 6, 7, 8$$

where the values represent the left, upper and right vertices. It can be seen that $r(3)$ is moved into the right. This generates more emphasis when values are higher, and it highlights situations with higher risk associated.

3) Rules set design: The rules set maps fuzzy inputs (i.e. after fuzzification) to the fuzzy output (i.e. before defuzzification). Traditionally, there is a rule defined for each possible combination of input variables. In our case, this would generate 2^{28} rules, which is not tractable computationally. Therefore we need to adapt the input to reduce the dimensionality of the rules (i.e. rules pruning).

The input data is collected as time series and the model captures which meal is over the threshold. But this is not important for a clinician, as they focus on the percentage of values over the threshold. Thus, we will reformulate the input as

$$in = [post_{sorted}, pre_{sorted}]$$

where those vectors represent the input vector sorted descending.

This mapping (i.e. rules pruning) is equivalent to the complete system from the clinician perspective. Let's consider the fasting and postprandial values separately. For fasting, there are seven values. As they are sorted decreasingly, the possible combinations are: None of them is over the threshold, or only the first one, or only the first two, and so on, until all values are over the threshold. Therefore, there are 8 possible combinations. Using the same logic for postprandial, there are 22 different possible combinations. As they are independent, given one set of values for fasting, any set of values for postprandial is possible. This mapping reduces the number of rules to $22 \times 8 = 176$. For example, let's assume that there are 2 values over the threshold. Previously there were around 400 different combinations, whereas now all those combinations are mapped into a 3 different options (both in postprandial, both in fasting, one in each). We will abuse notation, and use the naming of 1 to refer to the sorted input vector from now on.

To define the rules, we assign an output membership value to intervals of values that are above the threshold. This means that, if there are three or less values over the threshold, we assign fuzzy risk $r(0)$, if there are less than five values we assign risk $r(1)$, and so on. Note that $r(3)$, which we shown shifted to the right, corresponds to the hard threshold currently used in the hospital (i.e. 30%). Exact values for this mapping depend on the implementation, and will be discussed in the section IV. Rules are associated a weight, corresponding to $w(r(i)) = 1/(i+1)$. The weight helps detecting samples with few values over the thresholds to reduce false positives. As shown in the current section, the proposed system is functionally simple, embeds domain expertise and rules from clinicians and, as we will see in the IV section, is flexibly able to provide a risk assessment in the considered medical procedure for GDM.

IV. Experiments and discussion

The experiments were executed on a single laptop, with an Intel(R) Core(TM) i7-8550U CPU @ 1.80 GHz 8 CPU, 16 GB RAM and Windows10 OS. The Fuzzy Inference System implementation together with the experimental parameters can be found in the Python code using the Scikit-Fuzzy library implementation at https://github.com/csalort/fuzzy-gdm. Following the steps in the README file in the Github repository, one can generate the paper's figures.

A. Experiments

In order to augment the available data, we extended the dataset with artificially generated data, using the same distribution shown in figures 1 and 2 for postprandial and fasting glucose levels, respectively. 4000 sets of samples (each set consisting of 28 values) are generated using the extracted distributions, and 1000 sets are generated using $mean + 1$ to assess behaviour at larger thresholds, for a total of 5000 evaluation samples. The exact methodology is as follows:

1) Create the model using inputs, outputs, and rules explained in section III.
2) Generate the samples, detect how many elements in each sample are above the thresholds, sort them according to the procedure explained in section III, calculate the risk level using the model, and store all the elements.
3) Plot the results.

Each element corresponds to one script in the source code.

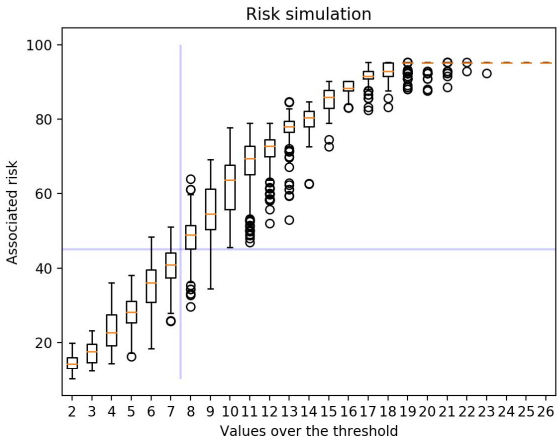

Fig. 7. Risk assessment simulation. The model assigns values ranging in the whole domain for different situations.

1) Evaluation and discussion: In figure 7 we can see the results of the experiments. Data is grouped along the X axis by the number of elements over the hard threshold. This is equivalent to the clinician feedback used currently in the hospital. The Y axis presents the risk level associated to each observation. The reference lines are drawn at 45 points, and

8 values, which are used as reference values for diagnosis. A detailed view can be seen in table I, where we show the metrics for calculating accuracy (true/false positives/negatives). The accuracy of the system (defined as (TP + TN) / n, where TP is true positives, TN is true negatives, and n is the number of experiments) is 94.68 % which represents that both criteria agree in most cases, and all the measurements with low or high risk are reflected in the risk score. Such insights in the data emerged from the inference and rule evaluation the the Fuzzy Inference System performs exploiting domain knowledge.

TABLE I
ACCURACY METRICS FOR THE EXPERIMENTS

Clinician input	Risk < 45	Risk ≥ 45
No	0.1932	0.0198
Yes	0.0334	0.7536

2) Border cases: In this subsection we will discuss border cases, particularly the ones where the risk factor is above a threshold (i.e. 45 points in this case) but the medical criteria is not, and vice-versa, representing upper left and lower right quadrants in figure 7. We expect that observations with a higher assigned risk value map to observations with a higher mean and standard deviation. We collected two sets of samples:

$$s_{low} = \{in = [post, pre] \,|\, risk(in) \leq 45 \,\&\, vot(in) \geq 8\}$$

$$s_{high} = \{in = [post, pre] \,|\, risk(in) \geq 45 \,\&\, vot(in) < 8\}$$

where vot represents the number of values over the threshold. Both distributions are normal, and we ran a Kolmogorov-Smirnov test between the post low and post high, and between pre low and pre high. Both test result positive (p-values of $2.32e^{-5}$ and $6.18e^{-4}$ respectively). We also fit the curves, with mean values of $post_{low} = 6.28$, $post_{high} = 6.37$, $pre_{low} = 4.77$ and $pre_{high} = 4.85$. This analysis shows that both sets are sampled from different distributions, and the set associated with higher risk values has higher glucose measurements. The risk system adapts better to each observation, giving a greater risk to patients with higher glucose level on average.

V. CONCLUSIONS

In this paper we presented a work in progress to improve monitoring of gestational diabetes mellitus. By using a flexible and explainable Fuzzy Inference System, we proved that we obtain comparable results in risk assessment for extreme cases (low and high), and we achieve personalized diagnosis in border cases. Some benefits of our system include simplicity and interpretability, as its core computational base uses a simple set of rules, combining input variables, easy to explain in daily language. Our next steps will focus on collecting more data, and using it to improve the model, based on observations. We will also collect feedback from the medical team using the system, in order to improve the expert knowledge embedded into the model. Finally, we will present the infrastructure that allows the real-time monitoring when it reaches a proper level of maturity.

REFERENCES

[1] S. Y. Chu, W. M. Callaghan, S. Y. Kim, C. H. Schmid, J. Lau, L. J. England, and P. M. Dietz, "Maternal obesity and risk of gestational diabetes mellitus," *Diabetes care*, vol. 30, no. 8, pp. 2070–2076, 2007.

[2] L. Bellamy, J.-P. Casas, A. D. Hingorani, and D. Williams, "Type 2 diabetes mellitus after gestational diabetes: a systematic review and meta-analysis," *The Lancet*, vol. 373, no. 9677, pp. 1773–1779, 2009.

[3] M. Rigla, G. Garca-Sez, B. Pons, and M. E. Hernando, "Artificial intelligence methodologies and their application to diabetes," *Journal of Diabetes Science and Technology*, vol. 12, no. 2, pp. 303–310, 2018, pMID: 28539087.

[4] S. Rouhani and M. MirSharif, "Data mining approach for the early risk assessment of gestational diabetes mellitus," *International Journal of Knowledge Discovery in Bioinformatics (IJKDB)*, vol. 8, no. 1, pp. 1–11, 2018.

[5] L. A. Zadeh, "Fuzzy sets," *Information and control*, vol. 8, no. 3, pp. 338–353, 1965.

[6] J. Warren, G. Beliakov, and B. Van Der Zwaag, "Fuzzy logic in clinical practice decision support systems," in *Proceedings of the 33rd Annual Hawaii International Conference on System Sciences*. IEEE, 2000, pp. 10–pp.

[7] M. Okpor, "Prognostic diagnosis of gestational diabetes utilizing fuzzy classifier," *International Journal of Computer Science and Network Security*, vol. 15, no. 6, pp. 44–8, 2015.

[8] E. D. Übeyli, "Automatic diagnosis of diabetes using adaptive neuro-fuzzy inference systems," *Expert Systems*, vol. 27, no. 4, pp. 259–266, 2010.

[9] A. S. Lane, M. A. Mlynarczyk, M. de Veciana, L. M. Green, D. I. Baraki, and A. Z. Abuhamad, "Real-time continuous glucose monitoring in gestational diabetes: A randomized controlled trial," *American journal of perinatology*, 2019.

[10] N. Mendes, R. Tavares Ribeiro, and F. Serrano, "Beyond self-monitored plasma glucose and hba1c: the role of non-traditional glycaemic markers in gestational diabetes mellitus," *Journal of Obstetrics and Gynaecology*, vol. 38, no. 6, pp. 762–769, 2018.

[11] C. Mencar, G. Castellano, and A. M. Fanelli, "Interface optimality in fuzzy inference systems," *International journal of approximate reasoning*, vol. 41, no. 2, pp. 128–145, 2006.

[12] S. Guillaume, "Designing fuzzy inference systems from data: An interpretability-oriented review," *IEEE transactions on fuzzy systems*, vol. 9, no. 3, pp. 426–443, 2001.

[13] C. Mencar, C. Castiello, R. Cannone, and A. M. Fanelli, "Interpretability assessment of fuzzy knowledge bases: A cointension based approach," *International Journal of Approximate Reasoning*, vol. 52, no. 4, pp. 501–518, 2011.

[14] A. Holzinger, C. Biemann, C. S. Pattichis, and D. B. Kell, "What do we need to build explainable ai systems for the medical domain?" *arXiv preprint arXiv:1712.09923*, 2017.