MEDINFO 2010
C. Safran et al. (Eds.)
IOS Press, 2010

doi:10.3233/978-1-60750-588-4-1145

A Self-organizing map based morphological analysis of oral glucose tolerance test curves in women with gestational diabetes mellitus

Laura Gaetano[a], Giacomo Di Benedetto[a], Andrea Tura[b], Gabriella Balestra[c], Franco M. Montevecchi[a], Alexandra Kautzky-Willer[d], Giovanni Pacini[b], Umberto Morbiducci[a]

[a] *Dipartimento di Meccanica, Politecnico di Torino, Italy*
[b] *Institute of Biomedical Engineering, CNR, Padova, Italy*
[c] *Dipartimento di Elettronica, Politecnico di Torino, Italy*
[d] *Department of Internal Medicine III, University of Wien, Austria*

Abstract

Gestational diabetes mellitus (GDM) makes women at risk of type 2 diabetes during their life. In order to predict this later abnormal glucose intolerance, several antepartum and postpartum predictors have been identified. In this study we conjecture that future evolution is predictable from morphology of the oral glucose tolerance test (OGTT) curves at baseline. To test our hypothesis, as a first step we evaluated the association between the curve morphologies of normal and diabetic patient condition at baseline. In particular, we analysed glucose and insulin curves of a group of women with a history of GDM. A Self-organizing map (SOM) was proposed to evaluate shape differences among control, normal, impaired glucose tolerance and diabetic curves shape. We compared our results with the currently applied clinical classification. We found that morphology contains information about the current status of the patient, because the SOM analysis clearly allows to discriminate subjects belonging to healthy or diabetic group. Moreover, SOMs highlighted additional information that could be used for prognostic purposes.

Keywords:

Gestational Diabetes Mellitus, Self-Organizing Map, Morphological analysis, Oral Glucose Tolerance Test

Introduction

Gestational diabetes mellitus (GDM) is defined as the diabetic condition with onset during pregnancy [1]. The prevalence of GDM seems to be proportional to the prevalence of type 2 diabetes; however, in general, GDM prevalence ranges from 1% to 14% of all pregnancies depending on the population studied [2]. In most cases, the onset of GDM is characterized by the symptoms of type 2 diabetes, i.e., an increased insulin resistance, and a decline in insulin secretion [3,4]. For this reason, gestational diabetes is often considered to be type 2 diabetes unmasked by pregnancy, also because they share common risk factors, such as obesity. In general, shortly after delivery, glucose homoeostasis is restored to the antepartum condition, but women with a history of GDM often show high blood pressure, atherogenic lipid profiles [5,6], and have a high risk of developing type 2 diabetes [7]. A systematic review showed that the incidence of diabetes among women with a history of GDM ranges from 3% to 65%, because of differences in the duration of the follow-up period and ethnicity [8]. This means that women who had gestational diabetes have at least a seven-fold increased risk of developing type 2 diabetes compared with those who had a normoglycaemic pregnancy [9]. Hence, women with a history of GDM represent a high-risk population and, since the risk of type 2 diabetes seems to be maintained for several years, they should attend regular assessments of their glucose tolerance condition. For these reasons, there is the need to develop appropriate preventive strategies and to identify reliable prognostic factors. In the last years, different antepartum and postpartum independent predictors of later abnormal glucose tolerance have been identified [10-14]. Several studies considered parameters or variables derived from the oral glucose tolerance test (OGTT) data as possible predictors of diabetes risk in the former GDM population, but a few studies considered the entire morphology of those curves.

In this study we hypothesize that the future evolution to a condition of normal glucose tolerance or type 2 diabetes is predictable from the morphology of the OGTT curves at baseline. In order to evaluate this potential predictor capacity, the first step is to evaluate if additional and useful information is contained into curves shape besides the two specific values used for current diagnosis of normal/diabetic condition. For this reason, we used a Self-organizing map (SOM), that is a neural network able to reduce high dimensional data into a low-dimensional topological map and display similarities, in order to cluster OGTT curves basing on their shape. Glucose and insulin curves derived from OGTTs of women with former GDM were considered. The SOM-based analysis was compared with the clinical, shape-independent classification of the glucose tolerance condition (i.e. the gold standard), in order to assess whether the morphology of the OGTT curves are (i) correlated to such condition, and (ii) maintain memory of the previous disease (GDM condition). Moreover, having a small number of curves at disposal, the other aim of this study is to find out an efficient method able to guide data mining in presence of small datasets.

Materials and Methods

Subjects

A group of 127 Caucasian women with GDM was investigated together with a control group (CNT) of 40 women without known risks for diabetes, and with normal glucose tolerance both during pregnancy and at the time of study. Gestational diabetes was diagnosed according to American Diabetes Association (ADA) criteria [15].

All women were recruited during pregnancy from the outpatient department of the University Clinic of Vienna, and gave written informed consent for participation in the study, which was approved by the local ethics committee. They were studied for a maximum of 5 years after delivery. All women underwent a standard 75-g OGTT every year. For each OGTT venous blood samples were collected at fasting and at 10, 20, 30, 60, 90, 120, 150 and 180 min afterward. In this preliminary analysis, however, only OGTT data at the baseline condition were used. According to the criteria proposed by ADA in 1997 [16], summarized in Table 1, the population was divided into a normotolerant group (NGT), a group with impaired glucose tolerance (IGT), and a group with type 2 diabetes (T2DM).

Table 1 – ADA classification of diabetes. G is the glucose concentration in blood measured at the beginning of the test and after 120 min

Glucose Values [mg/dL]	Clinical Classification
[G(0) < 100]	NGT
100 ≤ [G(0)] < 126 **OR** 140≤ [G(120)] <200	IGT
[G(0)] ≥ 126 **OR** [G(120)] ≥ 200	T2DM

As an example of the OGTT data, Figure 1 displays the measured mean glucose curves of NGT, IGT and T2DM subjects.

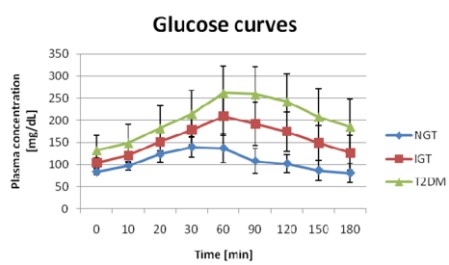

Figure 1 – Mean glucose curves and standard deviations of NGT, IGT and T2DM subjects

Morphological analysis

The Self-organizing map (SOM) is a subtype of artificial neural networks which uses a competitive learning technique to train itself in an unsupervised manner [17]. Competitive learning is an adaptive process in which the neurons of the neural network gradually become sensitive to different input categories, i.e., sets of samples in a specific domain of the input space. A division of neural nodes emerges in the network to represent different patterns of the inputs after training. The division is enforced by competition among the neurons: when an input x arrives, the neuron which is able to represent x better than the others wins the competition and is allowed to learn it even better. If there exist an ordering between the neurons, i.e. the neurons are located on a discrete lattice, the competitive learning algorithm can be generalized. Not only the winning neuron but also its neighbouring neurons on the lattice are allowed to learn: the whole effect is that the final map becomes an ordered map in the input space. This is the essence of the SOM algorithm. Technically, a SOM is made of m neurons located on a regular low-dimensional grid, usually one or two-dimensional. Each neuron i has a d-dimensional feature vector $w_i = [w_{i1},...,w_{id}]$, called prototype vector. At each training step t, a sample data vector $x(t)$ is randomly chosen for the training set, and the "distance" between $x(t)$ and each feature vector is computed. The winning neuron, denoted by BMU (Best Matching Unit), is the neuron with the feature vector closest to $x(t)$:

$$BMU = arg\min_i \| x(t) - w_i \|, \qquad i \in \{1,\ldots,m\}.$$

Once the winning neuron emerges, the weights of the neurons which are close to it are adjusted: because of the neighbourhood relationships, neighbouring neurons are pulled to the same direction, and thus prototype vectors of neighbouring neurons resemble each other. After the training is over, the map should be topologically ordered: this means that n topologically close input data vectors map to n adjacent map neurons or even to the same single neuron.

The morphological analysis was performed over all the glucose and insulin curves made available by the OGTT test. The analysis was conducted both on measured curves, and on curves obtained by the measured ones by removing their mean value. This was done in order to investigate whether curves can be someway classified exclusively in terms of their morphology, or the classification is biased by the exact value of each sample of the curve. For the SOM design, a hexagonal lattice map, a linear initialization of prototype vectors and a batch training algorithm was chosen, while the dimension of the grid depended on the size of the training sets used. The input sets were different, depending on what information we wanted to extract from the morphology of the considered curves. In detail, we investigated two different aspects:

- Curves belonging to CNT ($n = 40$) and NGT groups ($n = 43$) were used to construct a SOM with the aim of discovering whether a sort of memory of the previous disease (the GDM condition) is maintained in the morphology of the NGT curves (memory-of-disease-driven shape)
- Combination of curves belonging to CNT, NGT, IGT and T2DM groups (see Table 2) were used for training SOMs, aiming at evaluating whether there are substantial morphological differences among curves of subjects classified as having different glucose tolerance condition according to the ADA 1997 criteria [16].

Table 2 – Combination of curves used for training SOM

Combination of curves	Sample dimension
NGT vs. T2DM	($n_{NGT} = 43$) vs. ($n_{T2DM} = 37$)
NGT vs. IGT	($n_{NGT} = 43$) vs. ($n_{IGT} = 47$)
NGT vs. IGT vs. T2DM	($n_{NGT} = 43$) vs. ($n_{IGT} = 47$) vs. ($n_{T2DM} = 37$)

After training, the natural clustering tendency of curves was evaluated. Moreover, the available a priori knowledge about the input dataset was then used: each neuron was, in fact, afterwards labelled with class of the most numerous group of curves represented by that node. In this way, it was possible to understand how the current classification is represented by curves morphology and if SOM makes additional information come out.

The entire analysis was performed inside Matlab environment (The Mathworks, Inc., Natick, USA).

Results

First of all, there was any significant difference between results obtained with measured curves and those obtained with curves subtracted of their mean value (data not shown). For this reason, the following results are referred only to zero-mean curves.

A visual depiction of the analysis performed over glucose curves belonging to CNT and NGT groups is shown in Figure 2, where each colour is the result of the waveform classification performed by the SOM, i.e.: blue colour is associated to neurons representative of only CNT curves; yellow colour is associated to neurons representative of only NGT curves; green colour is associated to neurons with prototype vectors morphology is very close to both CNT and NGT curves. Notably, colours do not identify well separated regions in the map. This means that SOMs were not able to assign CNT and NGT curves to distinct regions of the map: neurons representative of only CNT curves were not confined in a particular region well separated from regions with neurons that recognized only NGT curves. Furthermore, about the 35% of prototype vectors was very similar both to CNT and NGT curves, and the corresponding neurons were quite randomly distributed in the SOM space. Also comparing different type of curves combined together (glucose and insulin) did not improve the SOM capability in discriminating among CNT and NGT curves (data not shown).

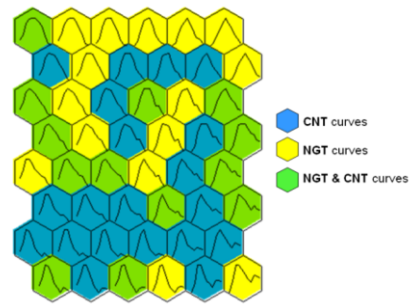

Figure 2 – CNT vs. NGT curves SOM: each neuron (hexagon) is characterized by its prototype vector (curve drawn inside the neuron). In this case, neurons representative of only CNT curves were not confined in a particular region well separated from regions with neurons that recognized only NGT curves

Concerning the morphological differences between NGT and T2DM groups, the SOM based analysis put in evidence a clear division in the morphology of the OGTT measured waveforms, as showed in Figure 3, in particular when only glucose curves were considered.

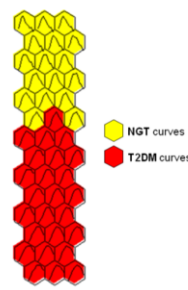

Figure 3 – NGT vs. T2DM curves SOM: upper region recognized only NGT patients, while lower area only T2DM curves

When the IGT group was included in the analysis, SOM analysis identified specific regions in the map referable to NGT, IGT and T2DM curves (Figure 4: glucose and insulin curves combined). However, when IGT curves were intro-duced in the analysis, the clear distinction observed between different group, such as Figure 3, was partially lost: transition regions appeared in the map because their prototype vectors were close enough to two different types of curve morphol-ogy. This is due to the fact that the inclusion of IGT curves in the analysis made less clear the distinction between NGT and T2DM curve morphology.

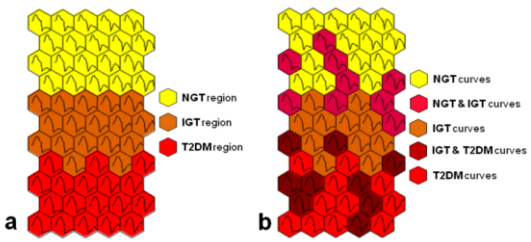

Figure 4 – NGT vs. IGT vs. T2DM curves SOM: it was possible to identify regions corresponding to different groups (a), however, some neurons had prototype vectors closed to two kind of curves, i.e. NGT and IGT or IGT and T2DM (b)

The distribution of the curves within neurons when the three clinically defined groups are included in the SOM analysis is summarized in Table 3. Notably, the number of NGT and T2DM curves falling into regions different from their clinical diagnosis was only the 2% and the 16%, respectively. In both the two cases, this happened because they were considered to belong to the same neurons as IGT curves.

Table 3 – Classification of curves using regions identified by SOM

Glucose tolerance condition	Areas identified by the SOM		
	NGT region	IGT region	T2DM region
NGT	42 *(98%)*	1*(2%)*	0
IGT	8 *(17%)*	30 *(64%)*	9 *(19%)*
T2DM	0	6 *(16%)*	31 *(84%)*

Discussion

To our knowledge, a few studies are present in the literature addressing the problem of the possible information contained in the shape of the OGTT curve [18]. In particular, in a population of women with a history of gestational diabetes, only a preliminary study [18] was carried out so far, based on a totally different approach from the SOM analysis used in this study.

Our results show that the whole morphology of the OGTT measured curves contain information about the current status of the patient with a history of GDM, because the SOM-based clustering clearly allows to discriminate subjects belonging to healthy or diabetic group even when the mean values is removed from the measured curves. Moreover, there are additional information that lead SOM to map nearer or not curves that currently belong to different groups. Exactly this topographic arrangement could be predictive of future evolution of patients.

Concerning the previous status of patient, OGTT measured curves of NGT women seem not to contain any memory of the previous GDM condition (that is, at this stage of analysis our method seems not to distinguish between NGT with former GDM and CNT). However, the other results presented here make us confident about the possibility of proceeding in our hypothesis verification: in fact, our preliminary findings show that OGTT curves morphology contains additional information about the current glucose tolerance condition that can be used for prediction of patient status evolution.

Conclusion

In conclusion, we succeeded in mining novel knowledge from our dataset even if it is relatively small: having not a large number of curves, through SOM, we have however extracted shape information that could be used for pattern recognition and feature selection in the next step, in which a relation between morphology characteristics and follow-up will be sought.

We also demonstrated the presence of additional information that, for example, leads some normal curve to be more similar to a diabetic one. The next phase will be to validate the hypothesis that these similarities or dissimilarities, recognized and underlined by SOM, are effectively prognostic factors of glucose tolerance evolution.

These results will be used to improve the follow up of women with a history of GDM. The fact that the whole morphology of the OGTT curves, and not only the absolute value of some glucose or insulin OGTT samples, depends upon the normal or diabetic condition of a subject allow for a more reliable identification of women that will develop type 2 diabetes.

References

[1] American Diabetes Association. Standards of medical care in diabetes - 2008. Diabetes Care 2008: 31 (1 Suppl): 12-54.

[2] American College of Obstetricians and Gynecologists Committee on Practice Bulletins - Obstetrics. ACOG Practice Bulletin. Clinical management guidelines for obstetrician-gynecologists. Number 30, September 2001 (replaces Technical Bulletin Number 200, December 1994). Gestational diabetes. Obstet Gynecol 2001: 98: 525-538.

[3] Kjos SL, Buchanan TA. Gestational diabetes mellitus. N Engl J Med 1999; 341 (23): 1749-56.

[4] Lencioni C, Volpe L, Miccoli R, Cuccuru I, Chatzianagnostou K, Ghio A, Benzi L, Bonadonna RC, Del Prato S, Di Cianni G. Early impairment of beta-cell function and insulin sensitivity characterizes normotolerant Caucasian women with previous gestational diabetes. Nutr Metab Cardiovasc Dis 2006: 16 (7): 485-93.

[5] Meyers-Seifer CH, Vohr BR. Lipid levels in former gestational diabetic mothers. Diabetes Care 1996: 19 (12): 1351-6.

[6] Pallardo F, Herranz L, Garcia-Ingelmo T, Grande C, Martin-Vaquero P, Jañez M, Gonzalez A. Early postpartum metabolic assessment in women with prior gestational diabetes. Diabetes Care 1999: 22 (7): 1053-8.

[7] Jovanovic L, Pettitt DJ. Gestational diabetes mellitus, JAMA 2001: 286 (20): 2516-8.

[8] Kim C, Newton KM, Knopp RH. Gestational diabetes and the incidence of type 2 diabetes: a systematic review. Diabetes Care 2002: 25 (10): 1862-8.

[9] Bellamy L, Casas JP, Hingorani AD, Williams D. Type 2 diabetes mellitus after gestational diabetes: a systematic review and meta-analysis. Lancet 2009: 373 (9677): 1773-9.

[10] Henry OA, Beischer NA. Long-term implications of gestational diabetes for the mother. Baillieres Clin Obstet Gynaecol 1991: 5 (2): 461-83.

[11] Metzger BE, Cho NH, Roston SM, Radvany R. Prepregnancy weight and antepartum insulin secretion predict glucose tolerance five years after gestational diabetes mellitus. Diabetes Care 1993: 16 (12): 1598-605.

[12] Damm P, Kühl C, Bertelsen A, Mølsted-Pedersen L. Predictive factors for the development of diabetes in women with previous gestational diabetes mellitus. Am J Obstet Gynecol 1992: 167 (3): 607-16.

[13] Coustan DR, Carpenter MW, O'Sullivan PS, Carr SR. Gestational diabetes: predictors of subsequent disordered glucose metabolism. Am J Obstet Gynecol 1993: 168 (4): 1139-44.

[14] Bian X, Gao P, Xiong X, Xu H, Qian M, Liu S. Risk factors for development of diabetes mellitus in women with a history of gestational diabetes mellitus. Chin Med J (Engl) 2000: 113 (8): 759-62.

[15] American Diabetes Association. Gestational diabetes mellitus. Diabetes Care 2003: 26 (1 Suppl): 103-5.

[16] The Expert Committee on the Diagnosis and Classification of Diabetes Mellitus. Report of the Expert Committee on the Diagnosis and Classification of Diabetes Mellitus. Diabetes Care 1997: 20 (7): 1183-97.

[17] Kohonen T. Self-Organizing Maps. 3rd ed. New York: Springer-Verlag, 2001.

[18] Tura A, Winhofer Y, Pacini G, Kautzky-Willer A. Shape index from OGTT data in former gestational diabetic women: new empirical parameter of beta-cell function? J Diabetes 2009, 1: A80.

Address for correspondence

Laura Gaetano
Dipartimento di Meccanica - Politecnico di Torino
Corso Duca degli Abruzzi, 24
10129 Torino (Italy)
E-mail: laura.gaetano@polito.it