

Predicción del nivel de glucosa en sangre para pacientes con diabetes utilizando técnicas evolutivas

J. Manuel Colmenar¹, Stephan M. Winkler² Gabriel Kronberger², Almudena Sánchez⁵, Esther Maqueda³, Marta Botella⁴, J. Ignacio Hidalgo⁵, Sergio Contador⁵, José Manuel Velasco⁵, Oscar Garnica⁵, and Juan Lanchares⁵

¹ Universidad Rey Juan Carlos, Madrid,
`josemanuel.colmenar@urjc.es`,

² University of Applied Sciences Upper Austria, Bioinformatics; Heuristic and Evolutionary Algorithms,
4232 Hagenberg, Austria
`stephan.winkler`

`@fh-hagenberg.at`, `gabriel.kronberger`
`@fh-hagenberg.at`

³ Endocrinología y Nutrición, Hospital Virgen de la Salud, Toledo,
`esthermaq@gmail.com`,

⁴ Endocrinología y Nutrición, Hospital Hospital U. Príncipe Asturias, Alcalá de Henares, `marta.botella@salud.madrid.org`,

⁵ Universidad Complutense, Madrid
`absys@ucm.es`, `hidalgo@dacya.ucm.es`, `scontador@ucm.es`, `mvelascc@ucm.es`,
`ogarnica@ucm.es`, `julandan@ucm.es`

Resumen La “International Diabetes Federation” estima que 415 millones de personas en el mundo están afectadas por algún tipo de diabetes. El control del nivel de glucosa en sangre es un aspecto crucial para evitar importantes complicaciones a largo plazo así como situaciones críticas en el día a día. Los diabéticos han de ser capaces de estimar la dosis de insulina que necesitan antes de una comida para mantener su glucosa en un nivel apropiado después de la misma. En este estudio vamos a comparar dos modelos predictivos de la interacción glucosa-insulina basados en técnicas evolutivas: Regresión simbólica basada en Programación Genética y caracterización de patrones diarios utilizando Gramáticas Evolutivas. Los algoritmos se han evaluado utilizando datos medidos con monitores continuos de glucosa aplicados a pacientes reales junto con la información de sus ingestas de carbohidratos y dosis de insulina. Estos modelos han sido validados en cuatro horizontes temporales diferentes: 30, 60, 90 y 120 minutos. Nuestros experimentos muestran correlaciones altamente significativas entre los datos reales y las estimaciones proporcionadas por ambos modelos.

Keywords: programación genética, gramática evolutiva, diabetes.

1. Introducción

La Diabetes Mellitus (DM) es una enfermedad que afecta a más de 415 millones de personas en el mundo según la “International Diabetes Federation” (IDF, [7]). Esta enfermedad se caracteriza por un incremento de la densidad de glucosa en sangre y está causada por un defecto en la secreción de insulina en el páncreas o en la acción de la insulina sobre las células. El primer caso, la Diabetes Mellitus Tipo 1 (DMT1) es el resultado de un proceso autoinmune que destruye las células productoras de insulina. El segundo caso, la Diabetes Mellitus Tipo 2 (DMT2) es la consecuencia del desarrollo de una resistencia a la insulina en las células del paciente. En ambos casos la glucosa no entra en las células permaneciendo en el flujo sanguíneo y produciendo hiperglucemia. La hiperglucemia mantenida produce daño celular en múltiples órganos por lo que pueden aparecer complicaciones crónicas en el individuo tales como ceguera, afecciones en el riñón y heridas en las extremidades que pueden requerir de una amputación [9]. Otra complicación de la Diabetes es el descenso de glucosa en sangre (hipoglucemia) que en casos extremos puede suponer la muerte del sujeto. La DMT1, que representa el 10 % de los pacientes con DM en el mundo, es la que ofrece mayores dificultades para un control glucémico óptimo, y en la que el tratamiento con insulina se torna más complejo.

Los pacientes con DMT1 necesitan inyectarse insulina subcutánea para controlar su glucemia de por vida. Predecir la respuesta glucémica en función de la ingesta, ejercicio y dosis de insulina administrada es una tarea muy compleja que puede ser ayudada por dispositivos capaces de medir el nivel de glucosa de forma continua. En la práctica clínica habitual se emplean glucómetros para la determinación de glucemia capilar. También existen en el mercado distintos tipos de sensores que nos dan mediciones de glucemia de manera discreta pero frecuente (entre 1 y 5 minutos), denominados Monitores Continuos de Glucosa (MCG).

En este estudio, hemos investigado la viabilidad de 2 técnicas evolutivas para predecir los niveles de glucosa en pacientes de TDM1. Primero, presentaremos un modelo predictivo basado en regresión simbólica desarrollado mediante programación genética. En segundo lugar, presentaremos un modelo basado en gramáticas evolutivas que busca caracterizar el comportamiento del paciente en los periodos posteriores a la ingesta de carbohidratos. Los algoritmos se han ajustado utilizando datos extraídos de monitores continuos de glucosa aplicados a pacientes reales junto con la información de sus ingestas de carbohidratos y dosis de insulina.

En la literatura encontramos diferentes enfoques para el desarrollo de modelos predictivos. Por ejemplo, en [11], encontramos un trabajo enfocado a la predicción de híper e hipoglucemias basado en series temporales autoregresivas. Esta técnica tiene el problema de ser muy lenta en adaptarse a los cambios dinámicos que puede sufrir el paciente. En [5], se presentan varios modelos que son promediados para obtener la predicción y en [3], se propone un sistema desarrollado a partir de modelos básicos. No obstante, en estos trabajos nos encontramos con que la bondad de los predictores decae rápidamente según aumentamos el

horizonte de predicción. Esta situación nos motivó a buscar técnicas más complejas que fuesen capaces de modelar mejor el comportamiento de los pacientes y proporcionar predicciones más fidedignas a un mayor plazo.

El resto del artículo está organizado de la siguiente manera: en la sección 2 hacemos una descripción formal de las dos técnicas empleadas, presentando los resultados experimentales de ambas en la sección 3. Finalmente, en la sección 4, discutimos las conclusiones de este estudio y cuáles son las vías futuras que pretendemos seguir.

2. Técnicas empleadas

El objetivo de esta investigación es desarrollar modelos matemáticos específicos para cada paciente capaces de predecir los niveles de glucosa en sangre (*gluc*) futuros utilizando los valores de glucosa pasados, las ingestas pasadas y futuras de carbohidratos (*ch*) así como las dosis pasadas y futuras de insulina (*ins*). De este modo, para un paciente concreto tenemos:

$$\begin{aligned} future_{gluc}(t) &= f(gluc(t - 2h \dots t), \\ ch(t - 2h \dots t + 2h), ins(t - 2h \dots t + 2h), t) \end{aligned} \quad (1)$$

En las siguientes subsecciones, se describe el funcionamiento de las dos técnicas empleadas en este estudio.

2.1. Identificación de modelos predictivos utilizando regresión simbólica

Entre otras posibilidades, hemos aplicado la regresión simbólica utilizando Programación Genética (PG) con selección estricta de descendencia tal y como es descrita en [2] y [12]. Construimos un subconjunto de funciones compuestas a partir de un conjunto de funciones presentadas en [2] (tanto aritméticas como lógicas). En la selección estricta, los individuos recién nacidos sobreviven sólo si demuestran ser mejores que sus dos padres. En la figura 1 podemos ver esta técnica visualmente.

Para este estudio, hemos empleado la versión de PG que proporciona el software HeuristicLab [8]. HeuristicLab es un sistema de análisis de técnicas de optimización evolutivas (además de numerosos algoritmos de aprendizaje máquina). En esencia, nuestra meta es identificar los modelos para las siguientes variables que se interrelacionan en la producción de los valores futuros de glucosa en el horizonte t :

$$gluc_{f30}(t) = gluc(t + 30min) \quad (2)$$

$$gluc_{f60}(t) = gluc(t + 60min) \quad (3)$$

$$gluc_{f90}(t) = gluc(t + 90min) \quad (4)$$

$$gluc_{f120}(t) = gluc(t + 120min) \quad (5)$$

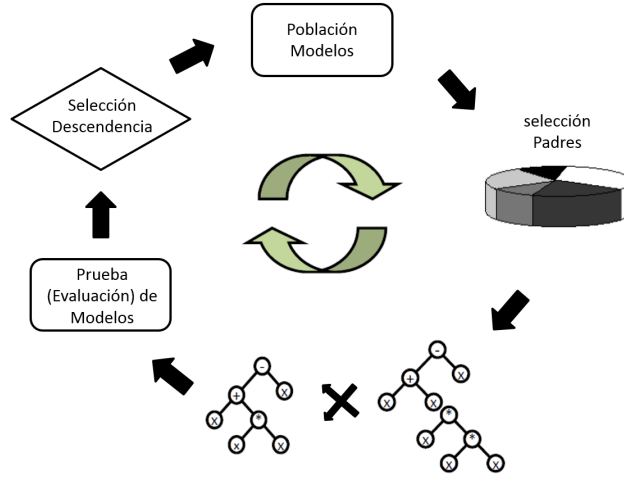


Figura 1. Programación Genética con selección estricta de descendencia [12].

Es decir, nuestro objetivo concreto es hallar modelos que describan los valores de glucosa en el horizonte t tal que:

$$gluc_{f30}(t) = f_{f30}(F(t)) \quad (6)$$

$$gluc_{f60}(t) = f_{f60}(F(t)) \quad (7)$$

$$gluc_{f90}(t) = f_{f90}(F(t)) \quad (8)$$

$$gluc_{f120}(t) = f_{f120}(F(t)) \quad (9)$$

2.2. Gramáticas Evolutivas

Se propone otro tipo de solución basada en Gramáticas Evolutivas (GE) [10], [4]. Las GE son un tipo de PG basada en gramáticas que combinan principios de biología molecular y el poder representativo de las formas gramaticales. La población en las GE está formada por un conjunto de individuos denominados cromosomas, cuyos fenotipos son obtenidos mediante un proceso de decodificación realizado por la gramática.

En el caso del modelado de la glucemia en pacientes diabéticos, el fenotipo de un individuo es la expresión del modelo que se utiliza para la predicción. Por tanto, hemos definido una gramática para que nos guíe en el proceso de optimización en la obtención de un modelo de predicción. La gramática definida considera que la predicción para un tiempo t depende de los valores de glucosa previos, de la cantidad de carbohidratos ingeridos y de la dosis de insulina.

Los algoritmos basados en las GE están programados en Java con la ayuda de la librería ABSys JECO [1]. Adicionalmente, hemos utilizado la aproximación

```

# Model expression
<func> ::= <exprgluc> + <exprch> - <exprins>

# Glucose
<exprgluc> ::= (<exprgluc> <op> <exprgluc>) | <preop> (<exprgluc>) | (<cte> <
    op> <exprgluc>)
    | predictedData(t-<idx
    >) | realData(t-<idx2hOrMore>)

# CH
<exprch> ::= (<exprch> <op> <exprch>)
    | <preop> (<exprch>)
    | (<cte> <op> <exprch>)
    | (getPrevData(1,t,1) * <cte> * <curvedCH>)

# Insulin:
## Sum of insulins in past 2h minus the peak
## Curve for the peak in past 2h
<exprins> ::= (<exprins> <op> <exprins>)
    | <preop> (<exprins>)
    | (<cte> <op> <exprins>)
    | getVariable(2,t-<idx>)

<op> ::= +|-|*|/
<preop> ::= Math.exp|Math.sin|Math.cos|Math.log
<cte> ::= <base>*Math.pow(10,<sign><exponent>)
<base> ::= 1|2|3| .. |97|98|99
<exponent> ::= 1|2|3|4|5|6|8|9
<sign> ::= +|-
<idx> ::= <dgtNoZero>|<dgtNoZero><dgt>|<dgtNoZero><dgt><dgt>
<dgtNoZero> ::= 1|2|3|4|5|6|7|8|9
<dgt> ::= 0|1|2|3|4|5|6|7|8|9
<ZeroOneTwo> ::= 0|1|2

```

Figura 2. Parte de la gramática desarrollada para la creación de modelos de glucemia.

de *fenotipos compilables* [6], que nos ha permitido la inclusión de código Java ejecutable como parte del fenotipo, aumentando la velocidad de ejecución del algoritmo en la evaluación de los individuos.

La figura 2 muestra una parte de las reglas que se han definido en la gramática. La fase de entrenamiento representada en la figura 3 consiste en una selección aleatoria de 5 días para cada paciente. Después, se han ejecutado las GE 10 veces para cada uno de los intervalos seleccionados. Por tanto, se obtienen 50 modelos para cada paciente para cada una de las comidas y un total de 150 modelos por día.

Los valores de los parámetros del algoritmo utilizado en la fase de entrenamiento se seleccionaron después de un conjunto de experimentos preliminares. Se utilizaron 2500 generaciones con una población de 250 individuos. La probabilidad de cruce y mutación fue de 0.7 y 0.01 respectivamente.

La figura 4 muestra un ejemplo de entrenamiento donde se obtiene un modelo para el desayuno del *pacienteA* de 1 día. La figura representa la predicción para los últimos 24 datos y, en este caso, el modelo es muy exacto.

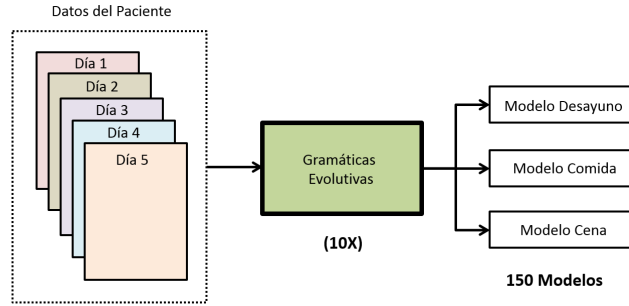


Figura 3. proceso de entrenamiento obtenido con GE para uno de los pacientes.

3. Resultados

3.1. Conjunto de Datos

Trabajamos con datos anonimizados de 3 pacientes reales. Los nombramos como *pacienteA*, *pacienteB* y *pacienteC*. Los 3 pacientes son mujeres con una edad media de 38 años. Los datos han sido adquiridos con un MCG y una bomba de insulina que el paciente llevaba sujeta al cuerpo. El periodo de muestreo para las 2 medidas fue de 5 minutos. También se anotó la cantidad de carbohidratos ingeridos por el paciente. Los datos se componen de al menos 10 días completos para cada paciente. Los días no son necesariamente consecutivos, ni tampoco los mismos días para los 3 pacientes.

3.2. Análisis

Con el objetivo de comparar los algoritmos, se ha calculado el coeficiente de correlación (ρ) entre los valores reales y los predichos. Tanto los valores ρ como los parámetros del modelo de cada paciente se calcularon por separado. Para el valor de la glucosa predichos en un tiempo t solo se utilizaron los valores de carbohidratos e insulina medidos después del tiempo t .

	$gluc_{f30}(t)$	$gluc_{f60}(t)$	$gluc_{f90}(t)$	$gluc_{f120}(t)$
<i>pacienteA</i>	0.921	0.783	0.638	0.504
<i>pacienteB</i>	0.913	0.790	0.681	0.555
<i>pacienteC</i>	0.902	0.778	0.669	0.586

Cuadro 1. Coeficiente de correlación normalizado (ρ) entre los valores medios reales y los estimados obtenidos con PG en diferentes instantes de tiempo: $t + 30$, $t + 60$, $t + 90$ y $t + 120$ minutos.

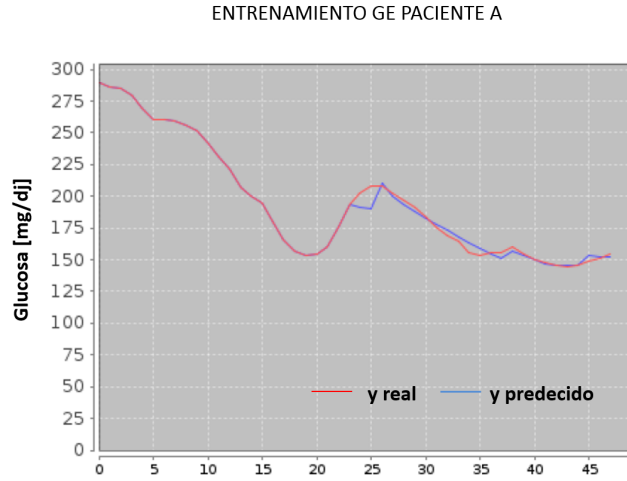


Figura 4. Evaluación de un modelo para el desayuno del *pacienteA* en la fase de entrenamiento con datos de 4 horas (48 valores en eje-x) utilizando GE. La glucosa se nombra como y .

La figura 5 muestra el gráfico de dispersión de las predicciones para el *pacienteA* realizadas con PG que estaban mejor correlacionadas. El cuadro 1 muestra los valores de ρ obtenidos para los instantes de $t + 30$, $t + 60$, $t + 90$ y $t + 120$ minutos. Se observa como los valores de ρ disminuyen a medida que el horizonte de predicción aumenta y por tanto el gráfico muestra una dispersión más grande.

En el caso de las GE, la fase de prueba se llevó a cabo de una manera diferente. Se seleccionaron aleatoriamente 2 días de datos para cada paciente. Para cada día, se utilizó el intervalo de tiempo correspondiente a una comida, y se evaluaron los modelos obtenidos en la fase de validación para cada paciente. Después, se calculó para cada tiempo t , los valores medios obtenidos por los

		Desayuno	Comida	Cena
<i>pacienteA</i>	Día 1	0.91	-0.827	0.182
	Día 2	0.945	0.399	-0.415
<i>pacienteB</i>	Día 1	0.513	0.883	0.97
	Día 2	-0.172	0.896	-0.962
<i>pacienteC</i>	Día 1	0.872	-0.658	0.961
	Día 2	0.45	0.086	0.949

Cuadro 2. Coeficiente de correlación normalizado (ρ) entre los valores medios reales y los estimados obtenidos con GE para las 3 comidas principales.

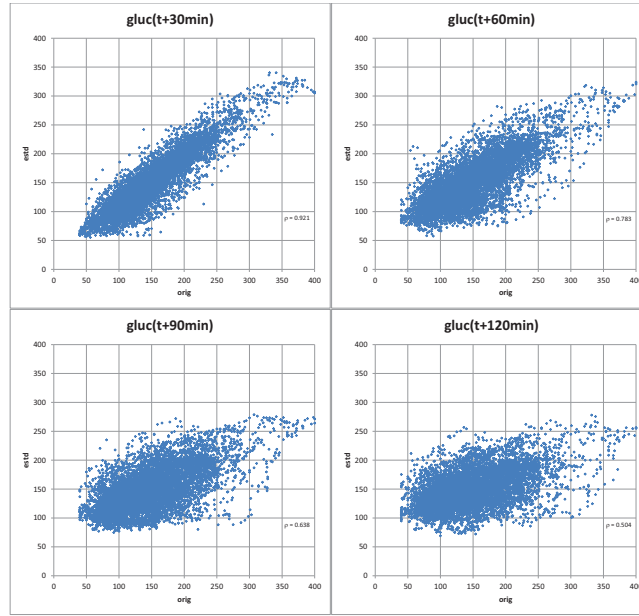


Figura 5. Revisión de los valores reales de glucosa vs valores futuros estimados para el *pacienteA*. Los valores estimados han sido obtenidos utilizando modelos de PG basados en identificación.

modelos que estaban contenidos en el rango $(0, 400]$. Este rango es establecido por los medidores actuales de glucosa. Por consiguiente, se considera que una predicción realizada fuera de ese rango no es válida.

Los valores de correlación para las predicciones se muestran en el cuadro 2, cada día por separado. Se puede observar que, en general, algunos modelos no son muy exactos. Correlaciones negativas significan relaciones inversas entre el valor real y el predicho (si una aumenta, el otro disminuye). Sin embargo, se obtuvieron varios valores (aparecen en negrita en el cuadro 2) reales y predicho muy parecidos entre sí.

	<i>PG</i>	<i>GE</i>		
<i>pacienteA</i>	0.504	0.9275	-0.214	-0.233
<i>pacienteB</i>	0.555	0.1705	0.8895	0.004
<i>pacienteC</i>	0.586	0.661	-0.286	0.955

Cuadro 3. Coeficiente de correlación normalizado (ρ) entre los valores medios reales y los estimados obtenidos con PG y GE. Horizonte de predicción = 120 minutos

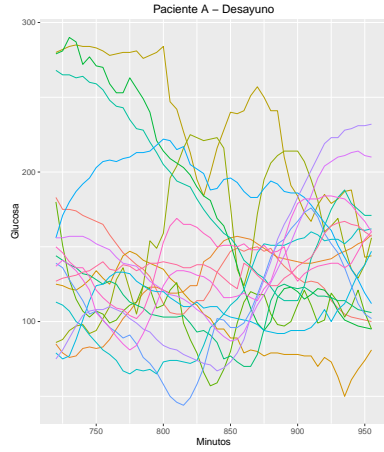


Figura 6. Paciente A - Niveles de glucosa en diferentes desayunos.

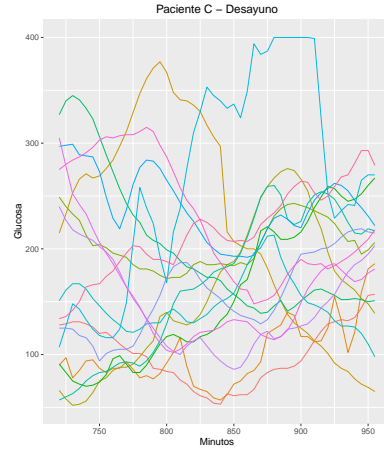


Figura 7. Paciente C - Niveles de glucosa en diferentes desayunos.

Para poder comparar las dos técnicas, se han utilizado para PG los valores ρ de la columna $gluc_{f120}(t)$ del cuadro 1 y para las GE los valores medios de los 2 días para cada periodo del cuadro 2. El cuadro 3 muestra los valores obtenidos. Se puede observar que los resultados de las GE tienen una varianza mucho más grande que los resultados de PG. Las GE es capaz de crear modelos con una correlación extraordinariamente alta para un determinado periodo y paciente pero al mismo tiempo puede incurrir en errores muy graves para el mismo paciente en un periodo diferente o para otro paciente en el mismo periodo. Sin embargo, los modelos obtenidos con la PG son más robustos, obteniendo valores ρ muy parecidos entre los pacientes a lo largo de todos los periodos.

Una explicación para esta situación la podemos encontrar en las figuras 6 y 7. En ellas se muestran el comportamiento de la glucosa durante varios días en el periodo del desayuno para los pacientes A y C (el comportamiento del paciente B es muy parecido al del paciente C). Podemos apreciar que resulta más fácil agrupar los diferentes días del paciente A que los días de los otros pacientes. Puesto que la técnica utilizada para las GE aprovecha la repetición de patrones para los diferentes periodos, podemos entender el acierto que consigue en ocasiones a costa de errores altos cuando los pacientes tienen comportamientos que no forman parte de los patrones habituales.

4. Conclusiones y Trabajo Futuro

En este artículo se han presentado 2 técnicas diferentes para la obtención de expresiones que modelen la glucemia sufrida por pacientes diabéticos. Con este objetivo, se ha trabajado con un conjunto de datos obtenidos de 3 pacientes, en concreto con los niveles de glucosa en sangre, cantidad de carbohidratos ingeridos y dosis de insulina recibida.

Se han aplicado técnicas de programación genética y gramáticas evolutivas para cada una de las comidas principales del día: desayuno, comida y cena. Los resultados obtenidos en la fase de prueba para algunos modelos fueron muy exactos, alcanzando valores de correlación entorno a 0,9.

Como trabajo futuro, se pretende avanzar en las estrategias propuestas, siendo este el comienzo de nuestro trabajo con datos obtenidos de pacientes reales. Además, nuestra intención es incorporar más variables de entrada a los modelos como el nivel de estrés del paciente o el número de horas de sueño.

Agradecimientos

Este trabajo ha sido financiado por los proyectos TIN2014-54806-R y TIN2015-65460-C2, así como por la Austrian Research Promotion Agency (FFG, K-project HOPL).

Referencias

1. Adaptive and Bioinspired Systems Group. ABSys JECO (Java Evolutionary COMputation) library. Available at: <https://github.com/ABSysGroup/jeco>, 2016.
2. Michael Affenzeller, Stephan Winkler, Stefan Wagner, and Andreas Beham. *Genetic Algorithms and Genetic Programming - Modern Concepts and Practical Applications*. Chapman & Hall / CRC, 2009.
3. Alain Bock, Grégory François, and Denis Gillet. A therapy parameter-based model for predicting blood glucose concentrations in patients with type 1 diabetes. *Computer methods and programs in biomedicine*, 118(2):107–123, 2015.
4. A. Brabazon, M. O'Neill, and I. Dempsey. An introduction to evolutionary computation in finance. *Computational Intelligence Magazine, IEEE*, 3(4):42–55, Nov. 2008.
5. C. Cobelli, C. Dalla Man, G. Sparacino, L. Magni, G. De Nicolao, and B.P. Kovatchev. Diabetes: Models, signals, and control. *Biomedical Engineering, IEEE Reviews in*, 2:54–96, 2009.
6. J. Manuel Colmenar, J. Ignacio Hidalgo, Juan Lanchares, Oscar Garnica, Jose-L. Risco, Iván Contreras, Almudena Sánchez, and J. Manuel Velasco. *Applications of Evolutionary Computation: 19th European Conference, EvoApplications 2016, Porto, Portugal, March 30 – April 1, 2016, Proceedings, Part II*, chapter Compilable Phenotypes: Speeding-Up the Evaluation of Glucose Models in Grammatical Evolution, pages 118–133. Springer International Publishing, Cham, 2016.
7. International Diabetes Foundation. IDF Diabetes Atlas 2015, <http://www.diabetesatlas.org/>.
8. Michael Kommenda, Gabriel Kronberger, Stefan Wagner, Stephan Winkler, and Michael Affenzeller. On the architecture and implementation of tree-based genetic programming in heuristiclabb. In *Proceedings of the 14th Annual Conference Companion on Genetic and Evolutionary Computation, GECCO '12*, pages 101–108, New York, NY, USA, 2012. ACM.
9. Roz D Lasker. The diabetes control and complications trial—implications for policy and practice. *New England Journal of Medicine*, 329(14):1035–1036, 1993.
10. Michael O'Neill and Conor Ryan. Grammatical evolution. *IEEE Trans. Evolutionary Computation*, 5(4):349–358, 2001.

11. G. Sparacino, F. Zanderigo, S. Corazza, A. Maran, A. Facchinetti, and C. Cobelli. Glucose concentration can be predicted ahead in time from continuous glucose monitoring sensor time-series. *Biomedical Engineering, IEEE Transactions on*, 54(5):931–937, may 2007.
12. Stephan M. Winkler. *Evolutionary System Identification: Modern Concepts and Practical Applications*. Schriften der Johannes Kepler Universität Linz. Universitätsverlag Rudolf Trauner, 2009.