



UNIVERSIDAD  
DE LOS ANDES  
VENEZUELA

PROYECTO DE GRADO

Presentado ante la ilustre UNIVERSIDAD DE LOS ANDES como requisito parcial para  
obtener el Título de INGENIERO DE SISTEMAS

[www.bdigital.ula.ve](http://www.bdigital.ula.ve)

ALGORITMO DE SELECCIÓN NEGATIVA PARA DETECCIÓN DE FALLAS EN  
EQUIPO DE INFUSIÓN DE INSULINA

Por

Br. Carmen Cecilia León Izarra  
Tutor: Rubén D. Rojas S.

Enero 2021

©2020 Universidad de Los Andes Mérida, Venezuela

C.C. Reconocimiento

# **ALGORITMO DE SELECCIÓN NEGATIVA PARA DETECCIÓN DE FALLAS EN EQUIPO DE INFUSIÓN DE INSULINA**

Br. Carmen Cecilia León Izarra

Proyecto de Grado — Control y Automatización, páginas

Escuela de Ingeniería de Sistemas, Universidad de Los Andes, 2020.

## **Resumen:**

La detección de fallas en equipos utilizados por pacientes que padecen diabetes tipo I, quienes requieren terapias continuas de infusión de insulina, es de vital importancia. Las fallas en los equipos de infusión de insulina (FEII) traen como consecuencia que la dosis de insulina requerida no sea administrada eficientemente, esto conlleva a la elevación de niveles de glucosa en sangre conocida hiperglucemia. La hiperglucemia prolongada está relacionada con complicaciones que pueden producir daños graves e irreversibles en los pacientes. Existen diversos enfoques utilizados en la de detección de fallas, entre los que se incluyen métodos basados en modelos cualitativos y cuantitativos del sistema y métodos basados en el historial de datos del sistema, que han permitido avances importantes en el área. Por ejemplo, el empleo de métodos estadísticos ha demostrado ser una herramienta con gran potencial en la detección de FEII. Sin embargo, siguen existiendo fallas que no son detectadas o no son detectadas antes de que las concentraciones de glucosa en plasma alcancen valores que provoquen complicaciones en los pacientes. Teniendo esto en cuenta, se puede considerar que un enfoque enmarcado en el campo de la computación inteligente basado en el sistema inmune biológico podría ser utilizado para lograr esta tarea. En particular, un método como el Algoritmo de Selección Negativa (ASN) que pertenece a los denominados Sistemas Inmunes Artificiales (SIA) y que ha sido ampliamente utilizado en aplicaciones de ingeniería relacionadas con reconocimiento de patrones y detección de fallas. En este trabajo se propone implementar el Algoritmo de Selección Negativa (ASN) usando datos provenientes de pacientes simulados que están recibiendo la dosis adecuada de insulina y datos provenientes de los mismos pacientes en momentos en los que se presentan fallas que afectan la administración de insulina para que, a partir de la clasificación de los mismos, se pueda detectar la FEII. Los datos son obtenidos del simulador UVA/PADOVA de pacientes con diabetes tipo I y clasificados por el ASN en dos categorías: valores normales y anormales, estos últimos se interpretan como la ocurrencia de alguna falla en el mecanismo de infusión

de insulina del paciente. Finalmente, se evalúa su desempeño al comparar los resultados obtenidos con investigaciones anteriores que emplean el enfoque de detección de fallas usando el análisis estadístico multivariable.

**Palabras clave:**

Diabetes tipo 1, fallas en equipo de infusión de insulina, detección de fallas, sistema inmune artificial, algoritmo de selección negativa.

[www.bdigital.ula.ve](http://www.bdigital.ula.ve)

# Índice

Capítulo 1	Introducción .....	1
<b>1.1 Antecedentes.....</b>	<b>3</b>	
<b>1.2 Planteamiento del Problema .....</b>	<b>8</b>	
<b>1.3 Objetivos.....</b>	<b>9</b>	
1.3.1 Objetivo General.....	9	
1.3.2 Objetivos Específicos.....	9	
<b>1.4 Justificación .....</b>	<b>10</b>	
<b>1.5 Metodología .....</b>	<b>11</b>	
<b>1.6 Alcance de la investigación.....</b>	<b>12</b>	
Capítulo 2	Marco Teórico.....	13
<b>2.1 Diabetes tipo I.....</b>	<b>13</b>	
2.1.1 Páncreas artificial en lazo cerrado.....	15	
2.1.2 Fallas en los Sistemas de Infusión de Insulina Continua.....	17	
<b>2.2 Simulador UVA/PADOVA.....</b>	<b>18</b>	
2.2.1. Escenario del simulador UVA/PADOVA .....	19	
2.2.2 Selección de los sujetos y mediciones de salida .....	21	
2.2.3 Plataforma de Pruebas.....	22	
2.2.4 Generación de los datos.....	23	
<b>2.3 Sistema Inmune Biológico (Nino, F y otros 2008).....</b>	<b>24</b>	
2.3.1 Discriminación entre lo propio y no propio .....	28	
2.3.2 Aspectos computacionales del sistema inmune .....	28	
<b>2.4 Sistema Inmune Artificial .....</b>	<b>30</b>	
2.4.1 Algoritmos basados en selección negativa .....	31	
2.4.2 Algoritmo de selección negativa de valores reales.....	33	
2.4.3 ASN con detectores de radio fijo .....	36	
2.4.4 ASN con detectores de radio variable.....	37	
Capítulo 3	Metodología .....	40

<b>3.1 Simulador UVA/PADOVA.....</b>	<b>40</b>
<b>3.2 Implementación del ASN .....</b>	<b>42</b>
3.2.1 ASN con radio de detección fijo .....	42
3.2.2 ASN con radio de detección variable .....	44
3.2.3 Validación del ASN .....	45
3.2.4 Ajuste de parámetros.....	46
3.2.4.1 ASN con radio de detección fijo .....	46
3.2.4.2 ASN con radio de detección variable .....	47
3.2.4.3 Desempeño del ASN utilizando nuevos parámetros .....	48
3.2.5 ASN en la detección de FEII .....	49
<b>3.3 Comparación entre ASN y Enfoque Estadístico Multivariable .....</b>	<b>52</b>
Capítulo 4     Análisis de resultados .....	53
<b>4.1 Simulador UVA/PADOVA.....</b>	<b>53</b>
<b>4.2 Implementación del ASN .....</b>	<b>55</b>
4.2.1 ASN con radio de detección fijo .....	55
4.2.2 ASN con radio de detección variable .....	56
4.2.3 Validación del ASN .....	58
4.2.4 Ajuste de parámetros.....	63
4.2.4.1 ASN con detectores de radio fijo .....	63
4.2.4.2 ASN con detectores de radio variable .....	65
4.2.4.3 Desempeño del ASN utilizando nuevos parámetros .....	67
4.2.5 ASN en la detección de FEII .....	70
<b>4.3 Comparación entre ASN y Enfoque Estadístico Multivariable .....</b>	<b>78</b>
Conclusiones .....	81
Referencias Bibliográficas.....	84

# Índice de Figuras

Figura 2. 1. Diagrama de bloques del control por realimentación aplicado en un páncreas artificial. _____	16
Figura 2. 2. Componentes principales del ambiente de simulación: modelo de glucosa-insulina (modelo humano), modelo del sensor, controlador y modelo de la bomba de insulina._____	19
Figura 2. 3. Interfaz de usuario del simulador UVA/PADOVA._____	21
Figura 2. 4 Selección de mediciones de salida y gráficos. _____	22
Figura 2. 5 Módulo correspondiente a la plataforma de pruebas del simulador UVA/PADOVA. _____	23
Figura 2. 6 Ejemplo de la estructura de los datos. _____	24
Figura 2. 7 Anatomía del sistema inmune (de Castro y Zuben, 1999). _____	26
Figura 2. 8 Patrón de reconocimiento en el sistema inmune. Izquierda: Célula B reconociendo un antígeno. Derecha: Célula T reconociendo un antígeno presentado por una célula accesoria. _____	27
Figura 4. 1 Degradación en rampa de la dosis de insulina para el paciente promedio simulado .....	54
Figura 4. 2 Respuesta de la glucosa en plasma para las tres fallas simuladas. Paciente simulado #4. ....	54
Figura 4. 3 Módulo correspondiente a la bomba de infusión de insulina del simulador UVA/PADOVA incluyendo la rampa de degradación de insulina.....	55
Figura 4. 4 Distribución del ancho y largo del sépalo. Conjunto de datos Iris. ....	58
Figura 4. 5 Distribución del ancho y largo del pétalo. Conjunto de datos Iris.....	58
Figura 4. 6 . Efectos de la variación del radio de detección sobre las tasas de detección y falsos positivos. ....	68
Figura 4. 7 Valores de glucosa en plasma paciente simulado #9 FEII 4:00 p.m. al utilizar ASN.....	73
Figura 4. 8 Valores de glucosa en plasma paciente simulado #4. Falla 4:00 p.m. ....	74
Figura 4. 9 Glucosa en plasma. Paciente simulado #3 Falla de las 4:00 p.m.....	75
Figura 4. 10Glucosa en plasma. Paciente simulado #1 Falla de las 12:00 p.m.....	75
Figura 4. 11Glucosa en plasma. Paciente simulado #2 Falla de las 12:00 p.m.....	76
Figura 4. 12 Glucosa en plasma. Paciente simulado #6. Falla del mediodía. ....	77
Figura 4. 13Glucosa en plasma paciente simulado #5. Falla de las 4:00 p.m.....	77

# Índice de Tablas

Tabla 4. 1 Tasa de detección obtenidas al aplicar el ASN de radio fijo utilizando conjunto de datos flores Iris.	59
Tabla 4. 2 Tasas de falsos positivos obtenidas al ejecutar el ASN de radio fijo, utilizando conjunto de datos flores Iris.	59
Tabla 4. 3 Tasas de detección obtenidas al ejecutar el ASN de radio variable. Utilizando conjunto de datos flores Iris.	60
Tabla 4. 4 Tasas de falsos positivos obtenidas al ejecutar el ASN de radio variable. Utilizando conjunto de datos flores Iris.	61
Tabla 4. 5 Número de detectores generados al ejecutar el ASN de radio variable. Utilizando conjunto de datos flores Iris.	62
Tabla 4. 6 Resultados obtenidos al ejecutar el ASN con detectores de radio fijo utilizando el conjunto Iris Virginica como conjunto propio. Variación del radio base de detección. Tasas de detección.	63
Tabla 4. 7 Resultados obtenidos al ejecutar el ASN con detectores de radio fijo utilizando el conjunto Iris Virginica como conjunto propio. Variación del radio de detección base (Rdb). Tasas de falsos positivos.	64
Tabla 4. 8 Resultados obtenidos al ejecutar el ASN con detectores de radio variable utilizando el conjunto Iris Virginica como conjunto propio. Variación del Rdb. Tasas de detección.	65
Tabla 4. 9 Resultados obtenidos al ejecutar el ASN con detectores de radio variable utilizando el conjunto Iris Virginica como conjunto propio. Variación del Rdb. Tasas de falsos positivos.	65
Tabla 4. 10 Resultados obtenidos ejecutando el ASN con detectores de radio variable utilizando el conjunto Iris Virginica como conjunto propio. Variación del porcentaje de solapamiento. Tasa de detección.	67
Tabla 4. 11 Resultados obtenidos ejecutando el ASN con detectores de radio variable utilizando el conjunto Iris Virginica como conjunto propio. Variación del porcentaje de solapamiento. Tasa de falsos positivos.	67
Tabla 4. 12 Resultados obtenidos aplicando ASN de radio fijo y variable. Utilizando conjunto de datos flores Iris y los parámetros calculados. Tasas de detección.	68
Tabla 4. 13 Resultados obtenidos aplicando ASN de radio fijo y variable. Utilizando conjunto de datos flores Iris y los parámetros calculados. Tasas de falsos positivos.	69
Tabla 4. 14 Resultados obtenidos aplicando ASN de radio fijo y variable. Utilizando conjunto de datos flores Iris y los parámetros calculados. Número de detectores.	70
Tabla 4. 15 Efectos de la variación del Rbd en las tasas de detección y falsos positivos.	71
Tabla 4. 16 Efectos de la variación del porcentaje de solapamiento en las tasas de detección y falsos positivos.	71

Tabla 4. 17 Tasas de detección porcentuales aplicando ASN para cada paciente simulado (%). Utilizando 75% del conjunto propio. _____	72
Tabla 4. 18 Tasas de falsos positivos porcentuales aplicando ASN para cada paciente simulado (%). _____	73
Tabla 4. 19 Tasa de falsos positivos (# fp/hora, Media (Desviación Estándar)). _____	78
Tabla 4. 20 Tiempo de detección (HH:MM, Media [Desviación Estándar])._____	79

[www.bdigital.ula.ve](http://www.bdigital.ula.ve)

## **Capítulo 1     Introducción**

La diabetes tipo I es un desorden metabólico provocado por un defecto en la secreción de insulina, lo cual conlleva a una elevación de los niveles de glucosa en sangre y alteraciones en el metabolismo de carbohidratos, grasas y proteínas. Ya que no existe cura para la diabetes tipo I, es necesaria la administración de insulina y el monitoreo de los niveles de glucosa en sangre de por vida. Es vital contar con regímenes efectivos de terapia de insulina que permitan minimizar las fluctuaciones de glucosa en sangre (valores de glucosa en sangre muy elevados o muy bajos, conocidas como hiperglicemia o hipoglicemia respectivamente). La terapia de insulina puede ser aplicada en inyecciones múltiples diarias o por infusión subcutánea continua. La infusión subcutánea continua de insulina implica la conexión, a través de un catéter, a una bomba de insulina que permita administrar la dosis requerida por un individuo (Fundación Mayo para la investigación y educación médica, 2019).

Un problema común encontrado en pacientes bajo terapia continua de insulina es la falla del equipo de infusión cuando el catéter de teflón o la aguja de acero se usan durante períodos de tiempo mayores a tres días. Las causas comunes de fallas en el equipo de infusión de insulina (FEII) incluyen equipos bloqueados, inflamación o fuga de insulina hacia la superficie de la piel. Estas fallas pueden hacer que la glucosa aumente a niveles

hiperglucémicos, incluso cuando el algoritmo de control (o el individuo) ha administrado la dosis de infusión de insulina requerida, ya que la misma no se está entregando de manera efectiva. La hiperglucemia prolongada se asocia con complicaciones a largo plazo en los ojos, riñones, nervios, corazón y vasos sanguíneos (Rojas, R., Garcia-Gabin, W., & Bequette, B. W., 2011). La importancia de contar con un sistema de detección de fallas eficiente en los sistemas de infusión continua de insulina que alerten al paciente que sus niveles de glucosa se están elevando, es evidente.

Una falla se puede definir como un alejamiento del rango aceptable de una variable observada o de un parámetro calculado asociado con un proceso. La detección de fallas es un área de gran importancia en procesos de ingeniería en general. Los métodos de detección de fallas se pueden clasificar en tres categorías generales: métodos basados en modelos cuantitativos, métodos basados en modelos cualitativos y métodos basados en el historial de proceso. Los enfoques basados en modelos se limitan a manejar modelos lineales y, en casos específicos, algunos modelos no lineales. Para un modelo no lineal en general, las aproximaciones lineales pueden ser deficientes y la efectividad de métodos de detección basados en estos modelos puede verse bastante reducida (Venkatasubramanian, Rengaswamy, Yin, Kvuri., 2003). El método seleccionado para el diseño de un sistema de detección y diagnóstico de fallas, se apoya en el campo de aplicación. Por ejemplo, en el área industrial, algunos enfoques aplicados incluyen métodos basados en el historial de proceso, específicamente, métodos estadísticos como el análisis de componente principal (ACP) (García A., Diego, 2009). El enfoque inmune, es un método basado en el historial de proceso,

ya que no requiere el manejo del modelo matemático del sistema para su utilización, sino que contando con un historial de datos que describan el comportamiento del sistema, es posible diseñar un mecanismo para detectar cualquier cambio que equivalga a un distanciamiento del comportamiento aceptable. Los algoritmos inmunes han sido ampliamente utilizados en la detección de fallas en distintos campos, ofreciendo importantes ventajas, tales como, elevadas tasas de detección, bajas tasas de falsos positivos y cortos tiempos de detección.

Tomando en cuenta que el sistema objeto de estudio, es un sistema fisiológico y que se dispone del historial de datos que describen el comportamiento del sistema (registro de datos correspondientes a los valores de glucosa en sangre de pacientes simulados que padecen diabetes tipo I), se considera pertinente utilizar un enfoque de detección basado en el historial de proceso para detectar FEII. Concretamente, se propone el uso de uno de los algoritmos inmunes más utilizados en detección de fallas, el Algoritmo de Selección Negativa (ASN).

## **1.1 Antecedentes**

El desarrollo de un páncreas artificial en lazo cerrado, el cual incluye un monitor continuo de glucosa en sangre, una bomba de infusión continua de insulina subcutánea y un algoritmo de control en lazo cerrado es objeto de una investigación continua. Esta tecnología ha sido revisada por Bequette (2005), Doyle (2007), Kumareswaran (2009), Cobelli (2009) y Bequette (2010). Estos estudios han permitido la mejora de los algoritmos de control del páncreas artificial cuyo desarrollo se ha acelerado en los últimos años gracias a la simulación

por computadores. En 2008 un modelo computacional fue aceptado por primera vez por la Administración de Alimentos y Medicamentos (FDA) de EE.UU. como un sustituto para pruebas preclínicas de ciertos tratamientos con insulina, incluyendo algoritmos en lazo cerrado (Kovatchev BP, Breton M, Dalla Man C, Cobelli, 2009). El primer simulador de diabetes tipo I, conocido como UVA/PADOVA, emuló los aspectos alimenticios e incluyó una población de 300 sujetos (100 adultos, 100 adolescentes y 100 niños). En el cada sujeto virtual es representado por un vector de parámetros extraído aleatoriamente de una distribución de parámetros adecuada (Visentin, R., Campos-Náñez, E., Schiavon, M., Lv, D., Vettoretti, M., Breton, M., & Cobelli, C. 2018).

Sin embargo, dichos dispositivos no se encuentran libres de fallas, en tal sentido, **Guilhem et al. (2006)** realizaron una investigación sobre los riesgos técnicos en la infusión subcutánea de insulina. En estudios utilizados por ellos como referencia, se reportan varios tipos de fallas en los equipos de infusión: obstrucción del equipo, fugas en el sitio de infusión, fugas en la conexión del equipo y fugas en el tubo de infusión. Los autores concluyen que aunque la terapia de infusión de insulina continua es eficaz en el tratamiento de pacientes diabéticos, los riesgos técnicos de esta terapia no deben ser subestimados.

**Kölle et al. (2018)** realizan una categorización de los métodos de detección de fallas y diagnóstico en sistemas de control de glucosa publicados con la finalidad de aportar una revisión del trabajo disponible a otros investigadores. Afirman que, frecuentemente, el primer paso para probar nuevos algoritmos de detección de fallas es realizar simulaciones en computadora. Sin embargo, los métodos publicados han sido probados en diferentes condiciones, incluyendo diferentes simuladores, lo que impide que los métodos puedan ser

comparados directamente unos con otros. Además, reportan que fallas como la interrupción en la infusión de insulina tienen efectos menos obvios y requieren métodos de detección complejos. Comparado con el número de estudios orientados a algoritmos de control de glucosa, existen pocas publicaciones relativas a detección y diagnóstico de fallas.

Una falla no detectada en el equipo de infusión de insulina puede conducir a una hiperglucemia y la cual de mantenerse a una cetoacidosis en unas horas. **Rojas, García y Bequette (2011)**, utilizan métodos estadísticos para detectar este tipo de fallas. Aplican clasificación bivariante, análisis de componentes principales (ACP) y un enfoque combinado. Sus estudios preliminares indican que el algoritmo de clasificación bivariante es sensible a cambios en la pendiente de glucosa en plasma, pero arroja una tasa de falsos positivos elevada. El segundo algoritmo propuesto, el ACP, permite obtener las menores tasas de falsos positivos, sin mayores cambios en el tiempo de detección. Por último, el enfoque combinado permite obtener menores tiempos de detección, pero la tasa de falsos positivos es elevada.

**Herrero et al. (2012)** proponen un sistema de detección de fallas robusto. Para probar el desempeño del sistema, utilizaron un simulador de pacientes con diabetes tipo I aceptado por la FDA. El algoritmo se basa en un modelo de intervalos para manejar las incertidumbres en los parámetros del modelo, medidas y entradas. De 100 pruebas realizadas, ocurrieron solo dos falsos negativos y un falso positivo. Aunque los intervalos de detección son largos (media de 200 minutos), las fallas son detectadas antes de que los niveles de glucosa plasmática alcancen un máximo de 300 mg/dl (límite de seguridad predefinido). La técnica propuesta permite detectar solo un tipo de falla, la cual consiste en la desconexión del

equipo de infusión de insulina, es decir, casos en los que ocurre una supresión total de la infusión de insulina. **Baysal et al. (2013)**, presentan enfoques que permiten detectar fallas en los equipos de infusión de insulina. Aplican un análisis estadístico multivariable que utiliza la pendiente del nivel de glucosa (aporte que reconocen a sus colegas Winston García y Ruben Rojas). En otro enfoque, proponen un umbral para el nivel de glucosa de 350 mg/dL. Para evitar exceso de falsos positivos, imponen un retraso de seis horas entre las alarmas de falla establecidas. Para evitar falsas alarmas, las alarmas se retienen hasta que el nivel de glucosa sea inferior a 250 mg/dL. Las medias de los tiempos de detección alcanzan 240 minutos.

**Cescon et al. (2016)**, proponen un algoritmo para la detección temprana de fallas. En el estudio participaron 23 sujetos, evaluados una vez a la semana durante 3 semanas. Debido a que el sensor de glucosa y el equipo de infusión fueron cambiados al mismo tiempo en la segunda semana del estudio, el algoritmo recibió inicialmente como entrada un flujo de datos proveniente de dispositivos mal calibrados, lo que afectó negativamente su desempeño. No se tomó ninguna medida para corregir este problema y lo dejaron pendiente para próximas investigaciones. El método se basa en tres componentes: segmentación de los datos, el cálculo de tendencia lineal del monitoreo de glucosa continuo y conteo de insulina. Los autores plantean la necesidad de realizar estudios clínicos a largo plazo para disminuir la tasa de falsos positivos. Los resultados, al evaluar el algoritmo retrospectivamente, corresponden a 50% de sensibilidad, 66% de especificidad y 55% de precisión.

Como parte de las investigaciones dirigidas a resolver problemas de ingeniería en general, es pertinente tomar en consideración los avances en biología y genética molecular por medio de los cuales se ha logrado comprender cómo funciona el sistema inmune. Esto

resulta de gran interés no solo desde el punto de vista biológico, sino también desde una perspectiva computacional. Así como el sistema nervioso inspiró el desarrollo de redes neuronales artificiales (RNA), el sistema inmune ha permitido el desarrollo de sistemas inmunes artificiales (SIA) como un nuevo campo de investigación de la inteligencia computacional (De Castro, L. N., & Timmis, J. I., 2002). De los métodos allí desarrollados, el algoritmo de selección negativa (ASN) es uno de los métodos más utilizados y las aplicaciones donde se utiliza con mayor frecuencia incluyen: detección de cambios, detección de fallas y detección de intrusión en redes (Li T. ,Computer immunology, 2004).

**Dasgupta y Forrest (1996,199)** presentan un algoritmo de detección de anomalías, para aplicaciones industriales, inspirado en mecanismos de selección negativa del sistema inmune, el cual discrimina entre lo propio y no propio. Definen propio a patrones de datos normales y no propio es cualquier desviación que excede una variación permitida

**Aguilar, Araujo y Aponte (2003)** plantean un sistema de detección de fallas en pozos de gas basado en el ASN. Concluyen que el modelo desarrollado puede ser utilizado en sistemas de alto riesgo y sistemas reales, donde es deseable detectar algún comportamiento anormal rápidamente.

**Dasgupta y Ji (2004)**, proponen una versión del ASN para datos de valores reales con detectores de tamaño variable. En su investigación proponen que un parámetro variable sea el tamaño de los detectores en un espacio de valores reales. Demostraron experimentalmente en su investigación que el esquema planteado es más efectivo, utilizando una menor cantidad de detectores, debido a su tamaño variable. Una de las ventajas que

ofrece esta versión del algoritmo es que el tiempo de ejecución, durante el proceso de generación de detectores y el proceso de detección, se reduce debido a la menor cantidad de detectores.

**Mohammadi et al. (2016)**, proponen un sistema de detección de intrusiones, basado en anomalías, en redes de computadoras utilizando SIA, capaz de reconocer nuevos ataques. Los resultados experimentales demostraron que los algoritmos propuestos tienen una respuesta rápida, bajas tasas de falsas alarmas, altas tasas de detección, y son rápidos al identificar nuevas muestras.

## 1.2 Planteamiento del Problema

Se han utilizado diferentes métodos y enfoques en la detección de fallas de equipos de infusión de insulina. Los resultados obtenidos en diversas investigaciones resaltan el potencial de los algoritmos desarrollados y apuntan a la necesidad de continuar realizando pruebas con el fin de perfeccionar su desempeño. En algunos casos, los algoritmos logran detectar solo fallas que implican una interrupción total de la infusión de insulina, en otros casos se reportan elevadas tasas de falsos positivos o bajas tasas de detección. El tiempo de detección también es un aspecto que requiere ser mejorado.

Partiendo del hecho de que se está realizando una investigación orientada a un sistema fisiológico y que se cuenta con el registro de datos que describen el comportamiento del sistema, se opta por proponer un enfoque basado en el historial de proceso para detectar

fallas en los equipos de infusión de insulina utilizados por estos pacientes. Es de gran importancia mencionar que la detección de la falla debe producirse en un tiempo que le permita al paciente tomar acciones antes de que la concentración de glucosa en plasma alcance valores que pongan en riesgo su vida. En este sentido, el presente trabajo de investigación propone el empleo de un enfoque basado en Sistemas Inmunes Artificiales (SIA). Específicamente, se propone emplear el Algoritmo de Selección Negativa (ASN) para determinar, si los valores de glucosa en sangre de un grupo de pacientes simulados están alcanzando niveles superiores a los normales, aún cuando el algoritmo de control del sistema está calculando la dosis de insulina adecuada que deberían estar recibiendo. Se busca comprobar si el ASN es capaz de mejorar el desempeño conseguido por otros métodos aplicados para la detección de fallas en este tipo de equipos.

### **1.3 Objetivos**

#### **1.3.1 Objetivo General**

Detectar fallas en equipos de infusión de insulina utilizado por pacientes con diabetes tipo I, en un tiempo que permita al paciente tomar acciones correctivas, utilizando el algoritmo de selección negativa (ASN).

#### **1.3.2 Objetivos Específicos**

- Estudiar el funcionamiento del simulador de pacientes diabéticos UVA/PADOVA.

- Estudiar las modificaciones realizadas a los modelos para simular fallas de equipos de infusión de insulina en el simulador de pacientes seleccionado, UVA/PADOVA.
- Estudiar la estructura de los conjuntos de datos generados a partir de simulaciones de las fallas que serán utilizados en diversas pruebas.
- Realizar la revisión bibliográfica de los algoritmos de los sistemas inmunes artificiales.
- Implementar el algoritmo de selección negativa según la literatura.
- Comparar los resultados, obtenidos empleando el ASN, con resultados obtenidos en otras investigaciones orientadas a la detección de fallas en equipos de infusión de insulina.

#### **1.4 Justificación**

Actualmente, existen investigaciones orientadas al desarrollo de algoritmos para la detección de fallas en equipos de infusión de insulina. Sin embargo, a pesar de haberse obtenido resultados favorables, en algunos casos, los algoritmos detectan solo fallas que implican la interrupción total de la infusión de insulina, más no casos de interrupción gradual. En otros casos, los algoritmos arrojan elevadas tasas de falsos positivos, bajas tasas de detección, o la detección no es realizada en un tiempo suficientemente corto como para evitar complicaciones en la salud del paciente.

Los Algoritmos de Selección Negativa (ASN) de los Sistemas Inmunes Artificiales (SIA), han sido ampliamente utilizados en casos de detección de fallas en aplicaciones

industriales, seguridad de redes, entre otras, demostrando que son una herramienta efectiva y con gran potencial en el área. El propósito de este trabajo de investigación es presentar un enfoque alternativo, basado en los SIA, para realizar la detección de fallas en equipos de infusión continua de insulina antes de que la concentración de glucosa en sangre alcance valores que pongan en riesgo la salud o la vida del paciente que lo utiliza.

## **1.5 Metodología**

En la primera etapa, se estudia el simulador de pacientes UVA/PADOVA. Se estudian las modificaciones que son necesarias hacer en el simulador que permitan obtener dos conjuntos de datos: un conjunto de datos de pacientes sometidos a terapia continua de infusión de insulina en momentos en los que el equipo de infusión no falla y un segundo conjunto de datos, que corresponde a datos de los mismos pacientes cuando se simulan fallas en el equipo de infusión. También se realiza un estudio de la estructura de los datos obtenidos, los cuales son utilizados como entrada al algoritmo detector de fallas.

Posteriormente, se realiza la revisión bibliográfica necesaria para obtener información relevante respecto a la diabetes tipo I, páncreas artificial, equipos de infusión de insulina y las fallas más comunes que se presentan en el uso de este tipo de equipos, así como también para conocer los conceptos básicos del sistema inmune. Se estudian los aspectos fundamentales de los SIA y algoritmos empleados en la detección de fallas, específicamente el algoritmo de selección negativa (ASN).

El ASN se implementa en MATLAB®. Luego, se valida utilizando para esto un conjunto de datos apropiado que permita evaluar su desempeño (tasa de falsos positivos y tasa de detección). Seguidamente, los datos generados con el simulador de pacientes UVA/PADOVA son utilizados como la entrada del ASN para determinar si es capaz de detectar las fallas simuladas en un tiempo pertinente para tomar acciones antes de que ocurran daños irreversibles al paciente. Finalmente, se comparan los resultados con resultados obtenidos en otras investigaciones en las cuales se utiliza un enfoque estadístico multivariable para la detección de fallas.

## **1.6 Alcance de la investigación**

Se propone aplicar un enfoque inmune para la detección de fallas en equipos de infusión de insulina utilizado por pacientes que padecen diabetes tipo I y reciben terapia continua de infusión de insulina. La detección debe producirse en un tiempo que permita al paciente tomar acciones antes de que su salud se vea afectada. Para tal fin, se utilizan datos provenientes del simulador de pacientes diabéticos UVA/PADOVA. Se implementan en MATLAB® dos versiones del ASN de los SIA y se utilizan los datos anteriormente mencionados como entradas a los algoritmos. Finalmente, se realiza una comparación con resultados obtenidos aplicando el enfoque estadístico multivariable. Debido a que en ambas investigaciones se utiliza el mismo simulador de pacientes y el mismo tipo de falla, la comparación es pertinente.

# **Capítulo 2**

## **Marco Teórico**

En este capítulo se presentan descripciones y conceptos teóricos relevantes para comprender qué son las Fallas de Equipos de Infusión de Insulina (FEII) y los Algoritmos de Selección Negativa (ASN) con los que se propone realizar la detección de las mismas.

### **2.1 Diabetes tipo I**

La diabetes tipo I, anteriormente conocida como diabetes juvenil o diabetes insulina dependiente, es una condición crónica en la cual el páncreas produce muy poca o no produce insulina. El papel de la insulina es transportar la glucosa desde el torrente sanguíneo hasta el músculo, grasa y otras células, donde puede almacenarse o utilizarse como fuente de energía. La glucosa (azúcar) es la fuente principal de energía para las células que forman los músculos y otros tejidos. La glucosa utilizada por el cuerpo proviene de dos fuentes principales: alimentación y el hígado. El azúcar se absorbe en el torrente sanguíneo, donde ingresa a las células con la ayuda de la insulina. El hígado almacena glucosa en forma de glucógeno. Cuando los niveles de glucosa están bajos, el hígado descompone el glucógeno almacenado en glucosa para mantener los niveles de glucosa dentro de un rango normal. En

la diabetes tipo I, como existe poca o no existe producción de insulina, la glucosa no puede entrar a las células y se acumula en el torrente sanguíneo. A pesar de que existe una investigación activa en este campo, hasta los momentos, la diabetes tipo I no tiene cura. El tratamiento se enfoca en manejar los niveles de azúcar en la sangre con insulina, dieta y estilo de vida que permitan prevenir complicaciones. La causa exacta de la diabetes tipo I es desconocida. Usualmente, el propio sistema inmune del cuerpo, el cual normalmente lucha contra elementos dañinos, erradamente destruye las células productoras de insulina en el páncreas. Una vez que un número significativo de células productoras de insulina del páncreas son destruidas, se produce muy poca o no se produce insulina (Fundación Mayo para la investigación y educación médica, 2019). Con el tiempo, las complicaciones de la diabetes tipo I pueden afectar órganos importantes en el cuerpo, incluyendo el corazón, vasos sanguíneos, nervios, ojos y riñones. Mantener un nivel normal de azúcar en la sangre puede reducir dramáticamente el riesgo de muchas complicaciones. Las complicaciones de la diabetes pueden ser incapacitantes e incluso potencialmente mortales (Rojas, R., Garcia-Gabin, W., & Bequette, B. W., 2011).

El tratamiento para la diabetes tipo I incluye: recibir insulina; conteo de carbohidratos, grasas y proteínas; monitoreo continuo de niveles de azúcar en sangre; alimentación saludable, ejercicio y mantenimiento de un peso saludable. La meta es mantener los niveles de azúcar en sangre los más cercano posible a valores normales para prevenir complicaciones. Generalmente, esto incluye niveles de azúcar durante el día antes de las comidas entre 80 y 130 mg/dL y, después de las comidas, valores no mayores a 180 mg/dL.

Las personas con diabetes tipo I necesitan terapia de insulina de por vida. Debido a que las enzimas digestivas destruyen la insulina, interrumriendo su acción, la insulina no puede ser administrada vía oral. Por lo tanto es necesario recibir insulina de una forma alternativa, para lo cual es posible utilizar uno de los siguientes mecanismos: inyecciones de insulina, bombas de infusión de insulina o una combinación de bombas de infusión de insulina, monitor de glucosa continuo y algoritmo de control en lazo cerrado (páncreas artificial en lazo cerrado). Un buen algoritmo de control para la administración de insulina en pacientes con diabetes tipo I debe incluir todas las variables, para así lograr normalizar los valores de glucosa en sangre. Después de prescribir una dosis inicial, la dosis debe ajustarse y basarse en el nivel de glucosa en la sangre. Este método de administración de insulina presenta un riesgo continuo de hiper e hipoglucemia porque las fluctuaciones de la glucosa momento a momento no se tratan adecuadamente con inyecciones de insulina subcutáneas intermitentes. El protocolo óptimo de administración de insulina es, por lo tanto, uno en el que el monitoreo de la glucosa en sangre y la dosificación de insulina se gestionen continuamente en tiempo real (Bequette, B. W., 2005).

### **2.1.1 Páncreas artificial en lazo cerrado**

Los tres componentes principales del sistema en lazo cerrado son: bomba de insulina continua, sensor de glucosa continuo y un controlador para ajustar la tasa de infusión de insulina basada en la señal de glucosa. Las principales limitaciones para desarrollar un páncreas artificial, son la disponibilidad de un sensor de glucosa robusto y confiable y un

algoritmo de control para regular glucosa en sangre bajo un amplio rango de escenarios de estados del paciente (Bequette, B. W., 2005).

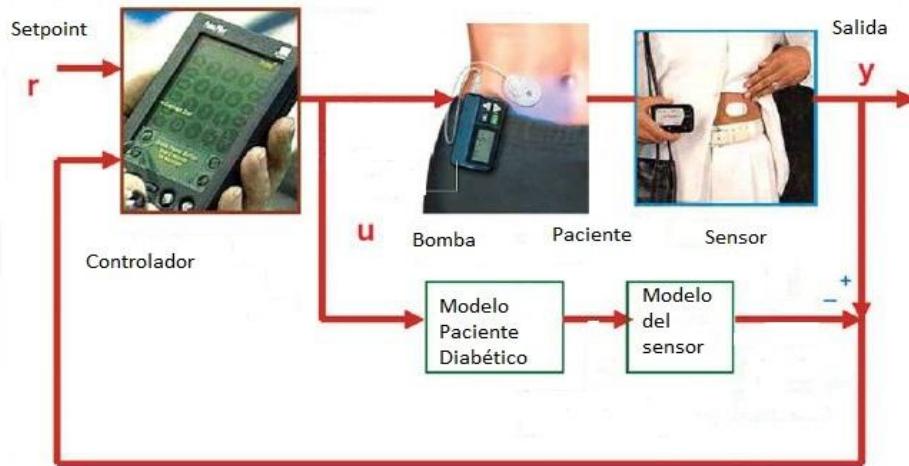


Figura 2. 1. Diagrama de bloques del control por realimentación aplicado en un páncreas artificial.

El principio básico del control por realimentación aplicado a un páncreas artificial se muestra en el diagrama de bloques de la figura 2.1. El valor de glucosa deseado es conocido como *setpoint*,  $r$ . La señal de salida (glucosa),  $y$ . El algoritmo de control está basado en la señal de error,  $e$ , la cual corresponde a la diferencia entre el *setpoint* y la señal de salida.

$$e = r - y$$

La señal manipulada (tasa de infusión de insulina) o salida del controlador es  $u$ . El controlador también puede tomar acciones de control si se utiliza una perturbación conocida para cambiar la entrada manipulada antes de que la perturbación afecte a la salida medida. Para el páncreas artificial, esta acción podría consistir en un bolo de insulina en anticipación a una comida, por ejemplo. Uno de los retos del control de glucosa en sangre es

el efecto de una comida, el cual puede ser visto como una perturbación al sistema. Algunas técnicas de control desarrolladas para el páncreas artificial, asumen que las perturbaciones ocurren como cambios en forma de escalón o rampa. Otro aspecto que se toma en cuenta para el desarrollo de los algoritmos de control, es la incertidumbre respecto a la cantidad de glucosa que es consumida en cada comida. También es importante mencionar que la sensibilidad a la insulina en cada individuo puede variar en períodos largos de tiempo, debido a cambios en su condición física y salud (Bequette, B. W., 2005).

Obviamente, el control de un páncreas artificial es mucho más que un simple algoritmo. El dispositivo debe ser fácil de programar y de usar para los especialistas médicos y para el paciente. Debe contar con una interfaz amigable, con un controlador fácil de configurar, el sensor debe ser fácil de calibrar y debe contar con un sistema de detección de fallas eficiente.

### **2.1.2 Fallas en los Sistemas de Infusión de Insulina Continua**

En los sistemas de infusión de insulina se reportan varios tipos de fallas, entre ellas, obstrucción del equipo de infusión, fugas en el sitio de infusión donde la aguja es colocada en el tejido subcutáneo, fugas en el lugar de conexión del equipo de infusión o fugas en el tubo de infusión. Generalmente, el sistema de alarma de los equipos de infusión no detecta las fugas. Más aún, en más de 85% de los eventos de oclusión, ocurre un deterioro metabólico antes de que se activen las alarmas de alta presión (Guilhem, I., Leguerrier, A. M., Lecordier, F., Poirier, J. Y., & Maugendre, D., 2006). Las complicaciones metabólicas

relacionadas con fallas en los equipos de infusión de insulina incluyen, cetoacidosis, hiper e hipoglicemia. La efectividad de las terapias de infusión continua de insulina subcutánea ha sido ampliamente probada en el control de glicemia en pacientes con diabetes tipo I. Por lo tanto, son mundialmente utilizadas en el tratamiento clínico de esta enfermedad. Sin embargo, los riesgos técnicos de esta terapia están presentes. De aquí la importancia de conducir investigaciones que permitan desarrollar mecanismos de detección de fallas eficientes.

## **2.2 Simulador UVA/PADOVA**

El simulador de pacientes diabéticos fue desarrollado por las Universidades de Virginia y Padova y aceptado por Administración de Alimentos y Drogas de los EE.UU, como sustitución de pruebas preclínicas para ciertos tratamientos con insulina, incluyendo algoritmos en lazo cerrado para el páncreas artificial (Visentin, R., Campos-Náñez, E., Schiavon, M., Lv, D., Vettoretti, M., Breton, M., & Cobelli, C., 2018). El simulador ha sido utilizado exitosamente por grupos de investigación académicos, así como también por compañías activas en el campo de la diabetes tipo I. El sistema ha demostrado que puede representar de forma adecuada los cambios de glucosa observados durante las comidas en pacientes con diabetes tipo I.

El simulador está implementado en MATLAB®. Incluye, entre otros módulos, un sensor para el monitoreo continuo de glucosa, el cual provee series de tiempo de observaciones consecutivas sobre el proceso de variaciones de glucosa en sangre. Otro

módulo importante es la bomba de insulina, la cual se utiliza para aproximar la entrega subcutánea de insulina tomando en cuenta el tiempo y dinámicas del transporte de insulina desde el tejido subcutáneo a la sangre. También cuenta con una interfaz de usuario que permite definir escenarios de prueba (horarios de comidas con las cantidades de carbohidratos correspondientes, etc.), permite seleccionar el grupo de individuos objeto de estudio y un conjunto de mediciones de salida (Visentin et al., 2018). En la figura 2.2 se presenta un esquema simplificado de los principales componentes del ambiente de simulación en el cual se basa el simulador UVA/PADOVA.

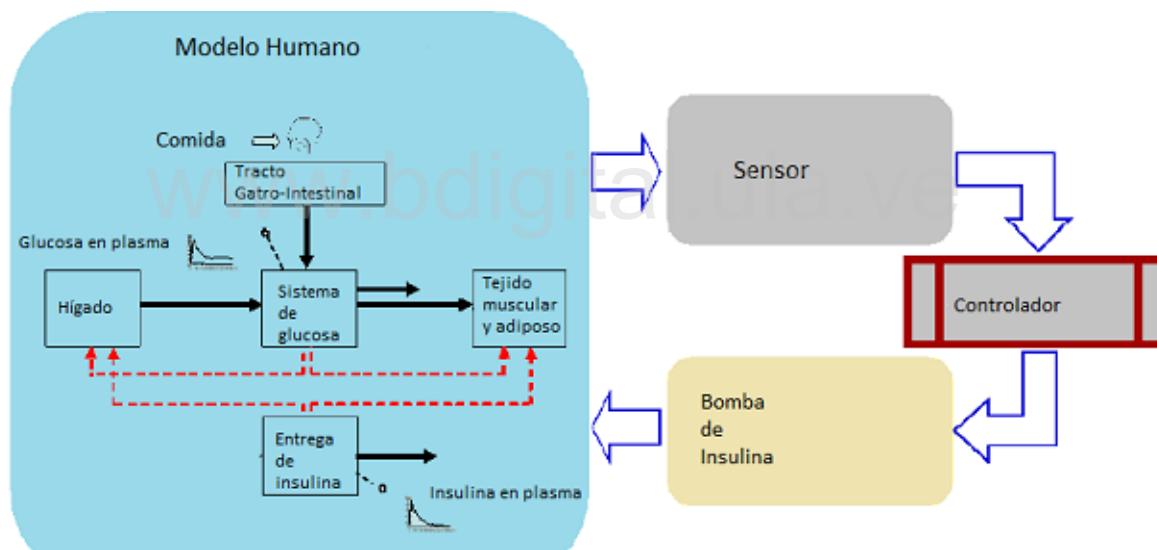


Figura 2. 2. Componentes principales del ambiente de simulación: modelo de glucosa-insulina (modelo humano), modelo del sensor, controlador y modelo de la bomba de insulina.

### 2.2.1. Escenario del simulador UVA/PADOVA

El simulador de pacientes diabéticos UVA/PADOVA cuenta con una interfaz de usuario que permite seleccionar el escenario en el cual se quieren hacer las simulaciones. Se incluyen tres opciones de escenario: escenario de una comida, escenario de un día y escenario de una semana. El escenario también puede ser creado en un archivo ASCII que puede ser cargado directamente desde la interfaz. El simulador carga el escenario

reconociendo algunas etiquetas en el archivo y transfiere la información para ser utilizada durante la simulación. En el escenario seleccionado para realizar las simulaciones, se incluye la siguiente información:

- Duración de la simulación.
- Unidades de tiempo de simulación.
- Tiempo del día en el que empieza la simulación (en minutos a partir de la medianoche).
- Tiempo de inicio del lazo cerrado.
- Unidad de tiempo del inicio del lazo cerrado.
- Un vector del tiempo de comidas.
- Unidad de tiempo del vector de comidas.
- Vector de la cantidad de comidas en gramos.
- Tiempo del bolo de insulina durante el lazo abierto.
- Unidad de tiempo del bolo de insulina.
- Cantidad del bolo de insulina (unidades de insulina) durante la etapa de lazo abierto.
- Tasa basal de insulina específica para cada paciente.
- Bolo óptimo, el cual es calculado en función de la cantidad de comida ingerida. Es específico para cada paciente.
- Inyección de insulina basal durante el lazo abierto.

En la figura 2.3 se muestra una imagen de la interfaz de usuario del simulador desde la cual puede ser directamente seleccionado el escenario en el cual se desea hacer las simulaciones, o en su defecto, desde donde puede ser cargado el archivo contentivo del escenario creado por el usuario.

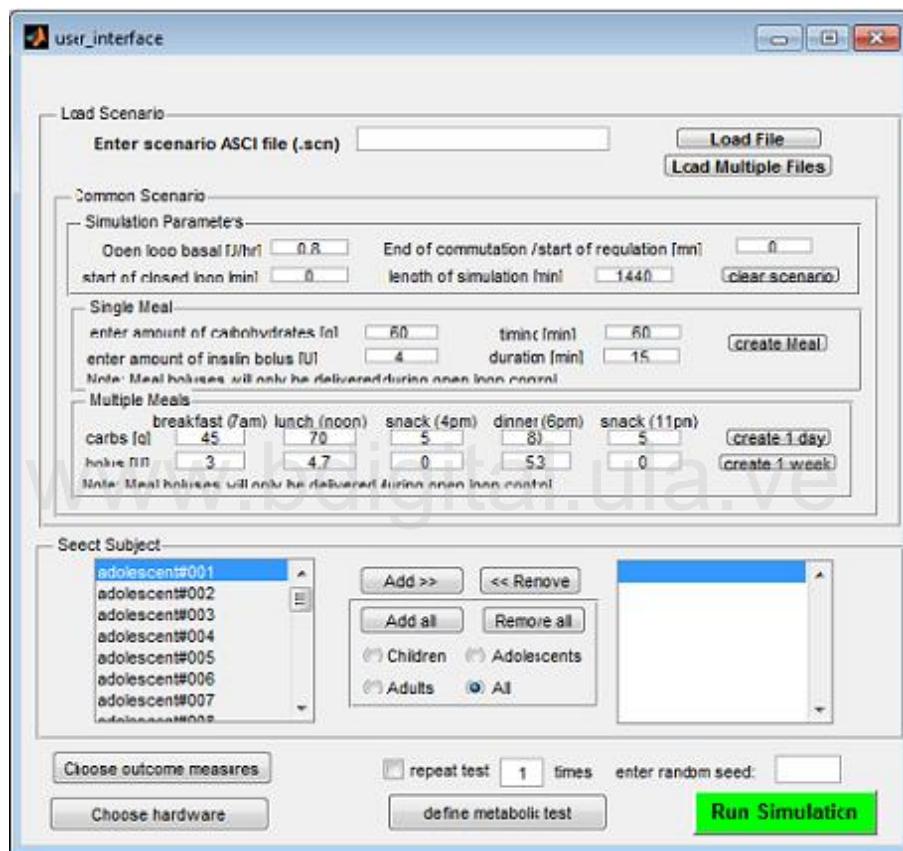


Figura 2. 3. Interfaz de usuario del simulador UVA/PADOVA.

## 2.2.2 Selección de los sujetos y mediciones de salida

El usuario puede seleccionar un grupo de sujetos (niños, adolescentes, adultos) agregando la categoría correspondiente en la interfaz de usuario de la figura 2.3. De igual manera, el simulador permite seleccionar las mediciones de salida y gráficas. El simulador

cuenta con ocho pantallas de gráficas y 15 mediciones. Las gráficas incluyen la simulación completa y las medidas de salida son calculadas en la fase de regulación. La siguiente figura corresponde a la ventana donde pueden ser seleccionadas las mediciones de salida.

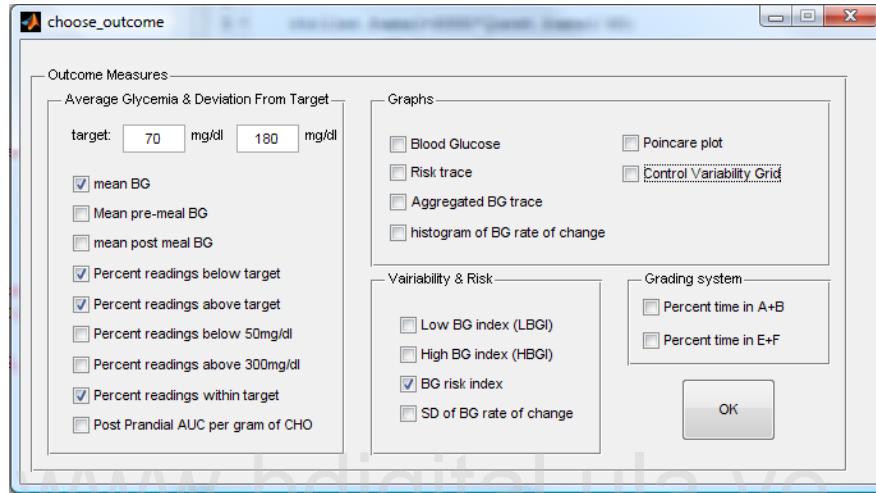


Figura 2. 4Selección de mediciones de salida y gráficos.

### 2.2.3 Plataforma de Pruebas

El archivo de plataforma de pruebas incluye el modelo de *Simulink* de la figura 4.5. El bloque color naranja contiene el controlador. En este modelo, el usuario puede reemplazar el bloque por su propio controlador y añadir el *script* de configuración desarrollado y cualquier función necesaria en la carpeta correspondiente del controlador. Si cualquier *script* o funciones son añadidas a una carpeta diferente, no será visible para el simulador y, por lo tanto, no podrá ser utilizada. De igual manera, es posible incluir nuevos bloques al modelo para simular, por ejemplo, algún tipo de falla.

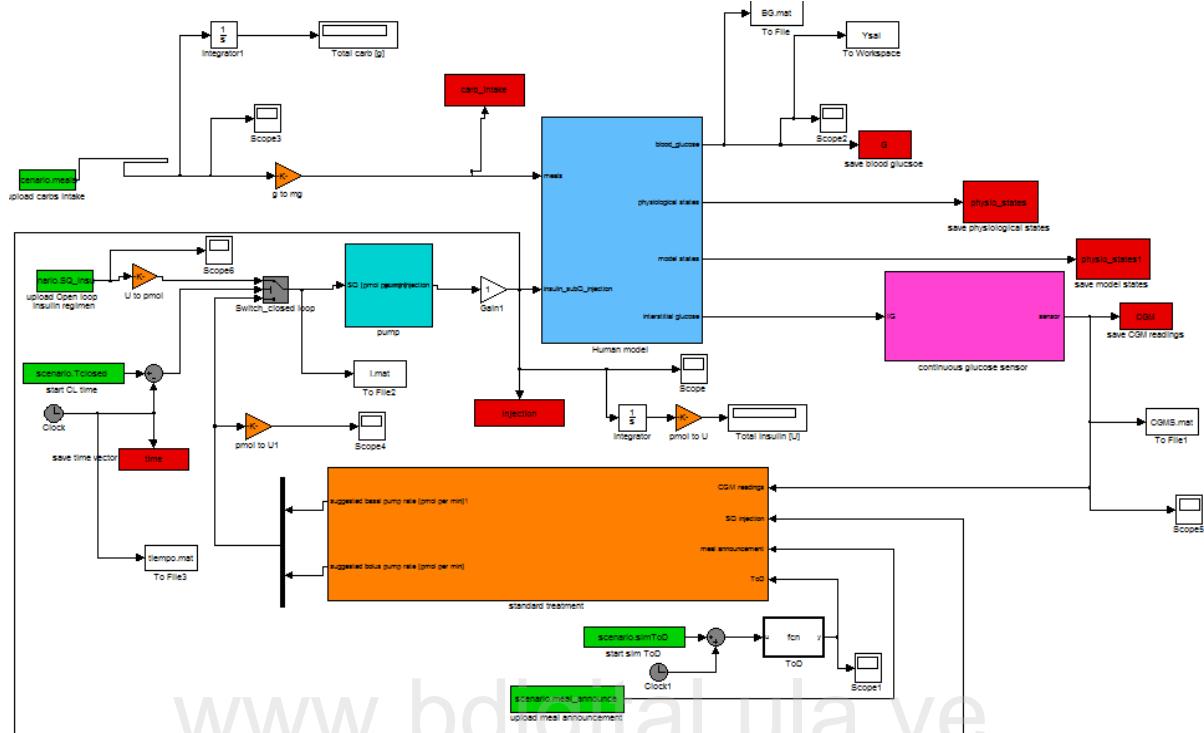


Figura 2. 5 Módulo correspondiente a la plataforma de pruebas del simulador UVA/PADOVA.

## 2.2.4 Generación de los datos

Los datos se almacenan en estructuras clásicas de MATLAB®. Para cada simulación la estructura de datos está compuesta por los resultados y el escenario. La estructura del escenario incluye los datos correspondientes a la configuración del escenario de pruebas: vector de cantidad de comidas en gramos, inyección de insulina basal durante el lazo abierto, duración de la simulación, tiempo en el que inicia el lazo cerrado, entre otras características específicas del escenario seleccionado. La estructura de los resultados incluye vector de valores de glucosa en sangre del paciente, vector de estados del sistema, vector de valores obtenidos del sensor de monitoreo continuo de glucosa, vector de dosis de inyección de

insulina, identificación del paciente y vector de tiempo de simulación. En la figura 2.6 se muestra un ejemplo de la estructura que tiene cada uno de los conjuntos de datos que fueron obtenidos con el simulador UVA/PADOVA.

```

Command Window
data =
    results: [1x1 struct]
    scenario: [1x1 struct]
>> data.scenario
ans =
    SQ_insulin: [1x1 struct]
        meals: [142x2 double]
    meal_announce: [1x1 struct]
        dose: [1x36 double]
        Tdose: [1x36 double]
        Qmeals: 0
    Tsimul: 10080
    Tclosed: 15000
    BGinit: []
        Treg: 0
        CR: 'off'
    simToD: 0
    basal: 0
    Obasal: 'quest'

fix >>

Command Window
>> data
data =
    results: [1x1 struct]
    scenario: [1x1 struct]
>> data.results
ans =
    G: [1x1 struct]
    states: [1x1 struct]
    time: [1x1 struct]
    injection: [1x1 struct]
    CGM: [1x1 struct]
    ID: 'child#002'

fix >>

```

Figura 2. 6 Ejemplo de la estructura de los datos.

### 2.3 Sistema Inmune Biológico (Nino, F y otros 2008)

El sistema inmune es una red compleja de tejidos, órganos, células y químicos especializados. Su función principal es reconocer la presencia de elementos extraños en el cuerpo y responder para eliminarlos o neutralizarlos. Todo ser vivo está expuesto a microorganismos (patógenos) que son capaces de causar enfermedades. Las sustancias que estimulan una respuesta inmunológica específica se conocen como antígenos, en otras palabras, los patógenos actúan como antígenos. El sistema inmune debe responder solo a

antígenos extraños, es decir, debe ser capaz de distinguir entre lo “propio” (células, proteínas y cualquier molécula que pertenece o es producida por el cuerpo) y lo “no propio”. En este sentido, las células que originalmente pertenecen a nuestro cuerpo y son inofensivas para su funcionamiento son llamadas propias, mientras que los elementos que causan enfermedades se denominan no propios. Una característica esencial del sistema inmune es el proceso de discriminación entre elementos propios y no propios, para lo cual cuenta con mecanismos sofisticados de reconocimiento de patrones y respuesta que utilizan una extensa red de mensajeros químicos. El sistema inmune es capaz de reconocer una variedad casi ilimitada de células infecciosas (elementos no propios).

El sistema inmune está compuesto por una gran variedad de moléculas, células y órganos que están esparcidos por todo el cuerpo. No existe un único órgano central controlando el funcionamiento del sistema inmune, sino que existen diferentes elementos transitando o en ubicaciones específicas que desempeñan roles complementarios. Los órganos que componen el sistema inmune se clasifican en órganos linfoideos centrales (médula ósea y timo) y órganos linfoideos periféricos (ganglios linfáticos, bazo, tejido mucoso del tracto respiratorio y digestivo). El propósito de los órganos linfoideos centrales es generar y ayudar en el proceso de maduración de las células linfoideas (linfocitos). Los órganos linfoideos periféricos facilitan la interacción entre los linfocitos y los antígenos.

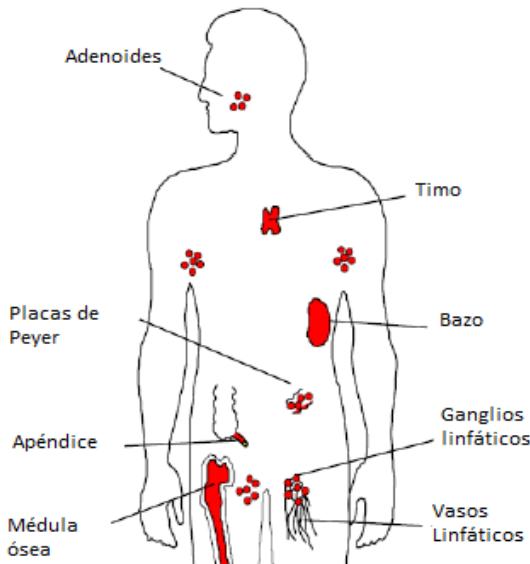


Figura 2. 7 Anatomía del sistema inmune (de Castro y Zubén, 1999).

Los linfocitos o glóbulos blancos son constituyentes importantes del sistema inmune. Estas células inmunes se producen en la médula ósea, circulan en la sangre y el sistema linfático, y residen en los órganos linfáticos para desempeñar funciones inmunológicas. La mayor población de linfocitos está constituida por linfocitos B y T. Los linfocitos T son células especializadas que maduran en el timo. Por otra parte, los linfocitos B son células especializadas capaces de reconocer antígenos particulares.

Desde la perspectiva del reconocimiento de patrones, una de las características más importantes del sistema inmune es la presencia de moléculas receptoras en la superficie de las células inmunes, capaces de reconocer una gama casi ilimitada de patrones antigenicos. Los dos tipos de células inmunes (linfocitos B y T) son bastante similares, pero difieren en cómo reconocen los antígenos y en sus roles funcionales. Las células B son capaces de

reconocer antígenos libres en solución (por ejemplo, en el torrente sanguíneo), mientras que las células T requieren que los antígenos sean presentados por otras células accesorias. En la figura 2.8 se observa que los antígenos están cubiertos con moléculas llamadas epítopes. Estas moléculas permiten que los antígenos sean reconocidos por los receptores (anticuerpos) ubicados en la superficie de los linfocitos B. Por otra parte, se observa que el antígeno debe ser procesado y presentado por una célula accesoria para que éste sea reconocido por los receptores de los linfocitos T.

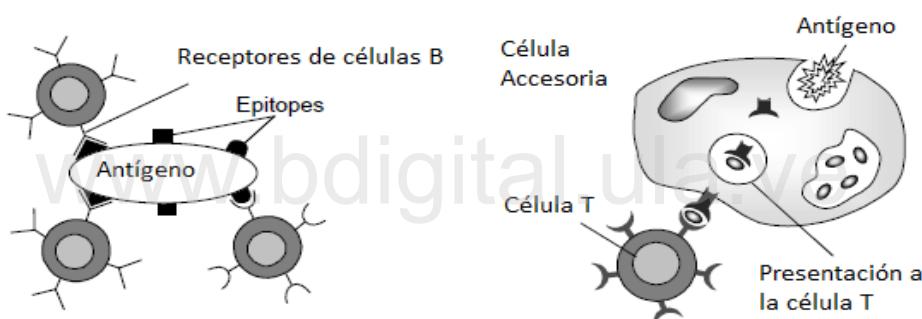


Figura 2. 8 Patrón de reconocimiento en el sistema inmune. Izquierda: Célula B reconociendo un antígeno. Derecha: Célula T reconociendo un antígeno presentado por una célula accesoria.

El reconocimiento de antígenos es el primer requisito para que el sistema inmune se active y se produzca una respuesta. El reconocimiento debe satisfacer ciertos criterios; primero, el receptor debe reconocer el antígeno con cierta afinidad, segundo, la unión entre el receptor y el antígeno debe ocurrir con una fuerza proporcional a esta afinidad. Si la afinidad es más grande que un umbral determinado (umbral de afinidad), entonces, el sistema inmune es activado.

Uno de los órganos linfoides centrales es el timo. El timo está ubicado detrás del esternón y desempeña un rol muy importante en la maduración de los linfocitos T. Una vez que los linfocitos T son generados, migran al timo para ser sometidos al proceso de maduración denominado selección negativa, en el cual, los linfocitos T que reconocen elementos propios son eliminados. Este proceso es la base del ASN implementado en los SIA y que ha sido ampliamente utilizado en distintos campos computacionales.

### **2.3.1 Discriminación entre lo propio y no propio**

Las células T maduran en el timo. Allí, atraviesan por un proceso de selección que asegura que son capaces de reconocer solo células no propias. Este proceso es denominado selección negativa. El propósito de la selección negativa es realizar pruebas de tolerancia a células propias. Las células T que reconocen células propias, fallan en esta prueba, y solo aquellas células T que no reconocen células propias son conservadas. Este proceso puede ser visto como un filtrado de la gran diversidad de células T. Las células receptoras de los linfocitos T que son capaces de discriminar entre propio/no propio, son llamadas detectores. Entre las principales funciones del sistema inmune biológico se distinguen reconocer y categorizar antígenos (Nino, F., & Dasgupta, D., 2008).

### **2.3.2 Aspectos computacionales del sistema inmune**

Desde un punto de vista del procesamiento de información, el sistema inmune biológico ofrece características interesantes, como por ejemplo:

- Coincidencia de patrones: el sistema inmune es capaz de reconocer antígenos específicos y generar respuestas apropiadas. Esto se logra a partir de un mecanismo

de reconocimiento que se basa en uniones químicas entre receptores y antígenos.

Estas uniones dependen de la forma de las moléculas y de la carga electrostática.

- Extracción de características: Generalmente, los receptores inmunes no se unen al antígeno completo, más bien lo hacen a porciones (péptidos) de los mismos. De esta manera, el sistema inmune puede reconocer un antígeno solo coincidiendo con segmentos de él. Los péptidos son presentados a los receptores de los linfocitos por las células presentadoras de antígenos. Estas células actúan como filtros que pueden extraer la información importante y remover el ruido molecular.
- Procesamiento distribuido: el sistema inmune no posee un controlador central. La detección y respuesta puede ser ejecutada local e inmediatamente sin comunicación con algún órgano central. Este comportamiento distribuido es logrado por billones de moléculas inmunes y células que circulan por los sistemas linfático y sanguíneo y son capaces de tomar decisiones en un ambiente de colaboración local.
- Aprendizaje y memoria: una característica importante del sistema inmune es su habilidad para aprender de la interacción con el ambiente. La primera vez que un antígeno es detectado, una respuesta primaria es inducida e incluye la proliferación de linfocitos. Algunos de estos linfocitos son conservados como células de memoria. La próxima vez que el mismo antígeno es detectado, las células de memoria generan una respuesta más rápida e intensa.

## 2.4 Sistema Inmune Artificial

Los sistemas inmunes artificiales (SIA) pueden ser vistos como abstracciones computacionales de los sistemas inmunes biológicos, por lo tanto, muchas de las técnicas utilizadas están basadas en los modelos teóricos del sistema inmune natural. Se han desarrollado algoritmos basados en los procesos de maduración y censado de las células T, principalmente en el proceso de selección negativa que permite la discriminación de lo propio/no propio. A partir de este proceso se desarrolla el Algoritmo de Selección Negativa (ASN) del cual existen diferentes variaciones. También, se han desarrollado algoritmos inspirados en la respuesta antigénica llevada a cabo por las células B, el algoritmo de selección clonal y redes inmunes artificiales. La teoría de selección clonal se utiliza para explicar la proliferación de células inmunes ante la presencia de un estímulo antigénico. Los algoritmos basados en la teoría de selección clonal son comúnmente utilizados en optimización y reconocimiento de patrones. La teoría de las redes inmunes fue desarrollada para explicar cómo se forma la memoria inmune, algunos de los campos de aplicación de los algoritmos basados en redes inmunes incluyen robótica y control. Se han explorado y aprovechado muchas de las características de los mecanismos SIA en distintas áreas de aplicación. En general, las técnicas de computación inmunológica han sido utilizadas para resolver problemas en un amplio rango de dominios tales como: optimización, clasificación, detección de anomalías, aprendizaje automático, control adaptativo, minería de datos, reconocimiento de imágenes y patrones, entre otros. Incluso han sido utilizados combinados con otros métodos como algoritmos genéticos, redes neuronales y lógica difusa. Aplicar un modelo basado en sistemas inmunes artificiales para resolver un problema

particular, requiere seleccionar el algoritmo inmune dependiendo del tipo de problema que se necesita resolver. Se deben identificar los elementos involucrados en el problema y cómo pueden ser modelados como entidades en un algoritmo inmune particular. Con base en sus características, algunas técnicas de los SIA se consideran más adecuadas para ciertas aplicaciones comparadas con otros enfoques. Se ha comprobado que los algoritmos de selección negativa han sido ampliamente usados en detección de fallas y seguridad computacional utilizando especialmente la propiedad de reconocimiento de lo propio y no propio (Nino, F., & Dasgupta, D., 2008).

#### **2.4.1 Algoritmos basados en selección negativa**

El algoritmo de selección negativa (ASN) consiste en dos fases: la etapa de generación y la etapa de detección. En la fase de generación, los detectores son creados aleatoriamente y sometidos a un proceso de prueba en el cual son eliminados si coinciden con alguna muestra propia (la finalidad es generar detectores que no coincidan con datos considerados propios y solo coincidan con datos que describen un comportamiento anormal del sistema). El criterio de coincidencia está basado en la representación de los datos. Luego de tener un número suficiente de detectores, determinado por el criterio de parada, la fase de generación termina. Los datos desconocidos, los cuales describen el comportamiento del sistema en estudio, son comparados uno a uno con el conjunto de detectores para ser clasificados como propios o no propios. Si un dato desconocido coincide con algún detector, es clasificado como no propio. Si el dato no es reconocido por algún detector, es asumido

como un miembro del conjunto propio (Nino, F., & Dasgupta, D., 2008) (Ji, Z., 2006). En la figura 2.9 se presenta un esquema general de las dos etapas del ASN.

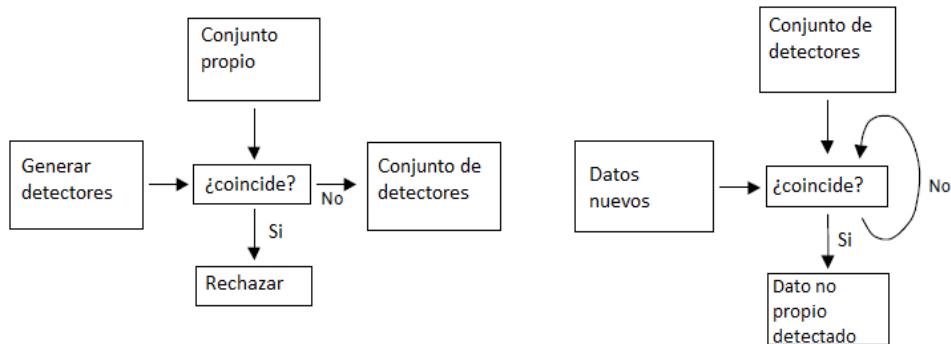


Figura 2. 9 Etapa de generación de detectores del ASN (izquierda). Etapa de detección (derecha).

El propósito fundamental del algoritmo de selección negativa es clasificar datos, por lo tanto el algoritmo es definido por el esquema de representación de los datos. Este estudio se enfoca en la representación de datos en valores reales. Para la representación de datos en vectores de valores reales, solo se han implementado esquemas de generación aleatoria de detectores. Otro aspecto importante en el desempeño del ASN, es la selección de la regla de coincidencia utilizada en el reconocimiento de datos. La selección de la regla de coincidencia o el umbral utilizado debe depender de la aplicación específica y de la representación de los datos. La regla de coincidencia es una medida de distancia, afinidad o similitud que dos datos comparten. Para una representación de los datos en vectores de valores reales, la regla de coincidencia común consiste en una medida de distancia matemática. El cálculo de una medida de distancia matemática arroja un número real que es asignado como afinidad,

permitiendo una comparación simplificada con el umbral de coincidencia asignado. La medida de distancia comúnmente implementada es la distancia Euclíadiana.

La representación de los datos y las reglas de coincidencia definen cada algoritmo de selección negativa, sin embargo, existen otros factores que afectan su desempeño. El número de detectores generado afecta la eficiencia de generación y detección y, en consecuencia, la velocidad del algoritmo. El criterio de parada y los esquemas de generación de detectores son parámetros de control típicos para determinar un adecuado número de detectores y cobertura (Ji, Z., 2006).

#### **2.4.2 Algoritmo de selección negativa de valores reales**

En el ASN de valores reales, por definición, los datos y los detectores son representados por datos de valor real (Nino, F., & Dasgupta, D., 2008). La representación del espacio propio/no propio corresponde a un subconjunto de  $\mathbb{R}^n$ . Se han desarrollado diferentes versiones del ASN de valores reales, las cuales incluyen:

- Algoritmo heurístico para generar detectores hiperesféricos.
- Selección negativa con reglas de detección (algoritmo para generar detectores hipercúbicos).
- Selección negativa de valores reales aleatoria (genera detectores utilizando un proceso aleatorio para optimizar la distribución de los detectores).
- Algoritmos con detectores de tamaño fijo y variable.
- Selección negativa con regla de detección difusa.

En los ASN de valores reales, un detector es definido como un vector n-dimensional que corresponde al centro y un valor real que representa su radio. La regla de coincidencia entre un detector y un dato anormal (antígeno), es expresada por una función miembro del detector, la cual es una función de la distancia entre el dato anormal y el detector. Existen distintas reglas de coincidencia (medidas de distancia), distintos mecanismos de generación de detectores y esquemas de detección que han sido implementados. Este estudio se enfoca en el empleo de la distancia Euclíadiana (Ecuación (2.1)), como medida de distancia y la generación de detectores de tamaño fijo y variable.

Distancia Euclíadiana:

$$\sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (2.1)$$

Los ASN de radio fijo y radio variable de valores reales tienen como objetivo principal generar un conjunto de detectores que cubran el espacio no propio. Por lo tanto, en el caso de ASN de valores reales, una buena aproximación del número de detectores necesarios para tal fin es obtenida a partir de la siguiente descripción:

$$N_{detectores} = \frac{V_{noPropio}}{V_d} \quad (2.2)$$

$V_{noPropio}$ : corresponde al volumen del espacio no propio.

$V_d$ : corresponde al volumen que ocupa un solo detector.

El volumen que ocupa un detector,  $V_d$ , se describe como el volumen de un hipercubo inscrito. Se toma esta definición porque es posible cubrir un espacio  $n$ -dimensional utilizando

hipercubos sin dejar espacios u hoyos. Por lo tanto, el volumen efectivo cubierto por un detector de radio  $r$  se calcula de la siguiente forma:

$$V_d = \left(\frac{2r}{\sqrt{n}}\right)^n \quad (2.3)$$

Para utilizar la ecuación (2.2) es necesario conocer el volumen del espacio no propio. Para este propósito, se parte del hecho de que el espacio total propio/no propio corresponde a un hipercubo unitario  $[0,1]^n$ . El volumen del espacio propio/no propio es igual a 1, por lo tanto, el volumen del espacio no propio se define como:

$$V_{noPropio} = 1 - V_{propio} \quad (2.4)$$

Un modelo para el espacio propio,  $\hat{S}$ , puede ser definido asumiendo que si un elemento está suficientemente cercano a un dato propio, puede ser considerado también como un dato propio. Esta cercanía se especifica por un umbral,  $u$ , que define la distancia mínima entre un elemento  $x$  y un dato propio, tal que este elemento  $x$  pueda ser considerado parte del conjunto propio, esto es:

$$\hat{S} := \{x \in U \mid \exists s \in S, \|s - x\| \leq u\}$$

Se define el volumen del espacio propio ( $V_{propio}$ ) como el volumen de  $\hat{S}$ , el cual se calcula como:

$$V_{propio} = V_{\hat{S}} = \int_U X_{\hat{S}}(x) dx$$

$X_{\hat{S}}(x)$ , corresponde a la función característica del conjunto  $\hat{S}$  definida por:

$$X_{\hat{S}}(x) = \begin{cases} 1 & \text{si } x \in \hat{S} \\ 0 & \text{si } x \notin \hat{S} \end{cases}$$

Utilizando muestreo aleatorio se puede estimar el volumen del conjunto propio,  $V_{\hat{S}}$ . El valor esperado de  $X_{\hat{S}}(x)$  es también una estimación de  $V_{\hat{S}}$ :

$$E[X_{\hat{S}}(x_i)] = \int_U X_{\hat{S}}(x) dx = V_{\hat{S}}$$

Ya que una buena estimación de la media de una variable aleatoria (valor esperado) es la media de un conjunto de muestras, se puede usar un promedio de  $\{X_{\hat{S}}(x_i)\}_{i=1\dots m}$  como un estimado del volumen del espacio propio:

$$V_{propio} \approx V_{\hat{S}} \approx \frac{\sum_{i=1}^m X_{\hat{S}}(x_i)}{m} \quad (2.5)$$

La estimación de la integral definida, utilizando el promedio de la variable aleatoria se conoce como la integración de Monte Carlo (Nino, F., & Dasgupta, D., 2008).

#### 2.4.3 ASN con detectores de radio fijo

El ASN consiste principalmente en dos fases. Primero, se genera un conjunto de detectores en la fase de entrenamiento o generación. Luego, en la fase de detección, los datos nuevos son examinados utilizando los detectores. En el ASN con detectores de tamaño constante, los detectores candidatos son generados aleatoriamente. Aquellos detectores que coinciden con los datos propios, utilizando como medida de coincidencia la distancia Euclíadiana, son eliminados. La fase de generación finaliza cuando el número preestablecido de detectores es generado. El radio de detección en este caso es el mismo radio del detector, el representa la distancia permitida al conjunto propio (Zhou, Ji, et al. 2004).

#### 2.4.4 ASN con detectores de radio variable.

En esta versión del ASN, se parte del hecho que el umbral utilizado como criterio de coincidencia define el radio de los detectores, y por lo tanto, es una buena opción hacerlo variable considerando que es probable que la región no propia cubierta por detectores sea de tamaño variable. La primera ventaja es que se puede cubrir mayor área del espacio no propio con menos detectores. Los pequeños espacios entre detectores y puntos propios no pueden llenarse con detectores de tamaño constante, sin embargo, utilizando detectores de tamaño variable, se pueden generar detectores de menor tamaño para cubrir los pequeños huecos, mientras detectores más grandes cubren el espacio no propio más amplio. Otra ventaja es que se puede utilizar la cobertura estimada como parámetro de control, en lugar de utilizar el número de detectores. Con el conjunto detector generado, el algoritmo puede automáticamente evaluar la cobertura estimada, proporcionando un criterio de parada útil. Primero, un conjunto de parámetros de control preestablecidos deben ser inicializados. El parámetro más influyente es el umbral o radio propio  $r_s$ , el cual constituye un importante mecanismo de balance entre la tasa de detección y la tasa de falsos positivos, en otras palabras, la sensibilidad y precisión del algoritmo. Debido a que los detectores no comparten el mismo radio fijo, se debe realizar una distinción entre el radio propio y el radio variable de los detectores  $r_d$ . Los otros dos parámetros de control que determinan el criterio de parada son: la cobertura estimada  $c_0$  y el número máximo de detectores  $D_{max}$ . La cobertura estimada es utilizada como uno de los criterios de parada, pero el parámetro que se toma en cuenta en la ejecución del algoritmo es  $m$ , el cual se relaciona con la cobertura estimada de la siguiente forma. Si se muestran  $m$  puntos en el espacio considerado y solo un punto no es

cubierto, asumiendo que 1 es la cobertura total, la región no cubierta estimada es  $\frac{1}{m}$ . Por lo tanto, la cobertura estimada es  $c_0 = 1 - \frac{1}{m}$ . Así, el número de intentos ( $m$ ) necesarios para asegurar la cobertura estimada,  $c_0$  es:

$$m = \frac{1}{1 - c_0} \quad (2.6)$$

La fase de generación del algoritmo comienza generando aleatoriamente detectores candidatos, pero en lugar de generar un conjunto completo determinado por un parámetro de control fijo, los detectores se generan uno a la vez. Cada candidato individual es comprobado utilizando el criterio de coincidencia determinada por la medida de distancia seleccionada. Si la distancia al punto propio más cercano es menor que el valor del radio propio, el detector es eliminado y se genera un nuevo candidato. Si la distancia mínima a cualquier punto propio es mayor que el radio propio, entonces el detector es almacenado temporalmente y la distancia es guardada como el radio  $r_d$ . Luego el detector candidato es chequeado para determinar si puede ser detectado por otro detector candidato almacenado previamente. Debido a que no es posible cubrir un volumen dado con detectores esféricos sin que exista solapamiento, se permite un porcentaje de solapamiento entre detectores. Si el detector coincide con otro, es eliminado, y el intento es almacenado en un contador que será utilizado para estimar la cobertura. Si el detector no coincide con algún detector previamente almacenado, es almacenado permanentemente para la fase de detección y el contador es llevado a cero. Si el número de detectores alcanza su número máximo, o el contador de intentos consecutivos que caen en puntos cubiertos alcanza un límite máximo (cobertura estimada), la fase de generación termina y en este último caso se estima con confianza que la cobertura es

suficiente para el espacio no propio. La fase de detección del algoritmo es similar a la del algoritmo de detectores de tamaño fijo. La única diferencia es que el umbral detector utilizado para la data desconocida se basa en el radio variable asignado a cada detector (Nino, F., & Dasgupta, D., 2008).

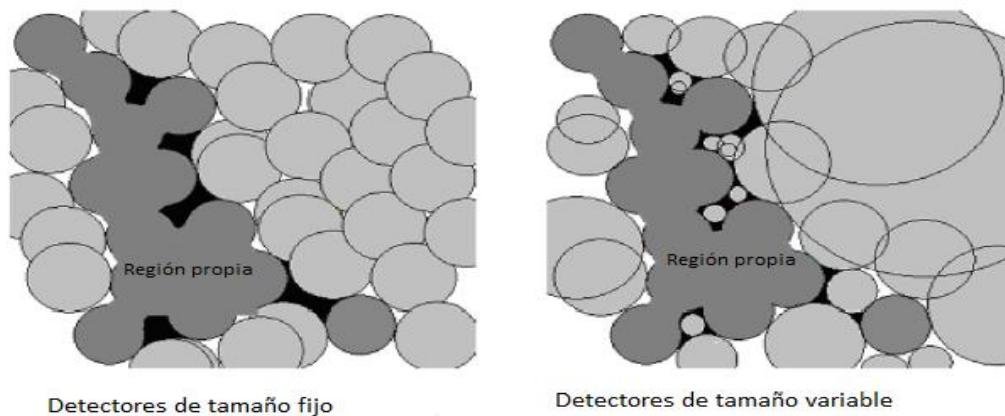


Figura 2. 11 Comparación de la cobertura de los detectores para los dos esquemas.

# **Capítulo 3**

## **Metodología**

En este capítulo se presenta la metodología aplicada para el desarrollo del mecanismo de detección de Fallas en Equipos de Infusión de Insulina (FEII) utilizando el enfoque inmunológico. En primer lugar, se presentan consideraciones con respecto al simulador de pacientes diabéticos UVA/PADOVA en el cual se simulan las FEII y se obtiene el conjunto de datos objeto de estudio. También, se explica el proceso de generación y la estructura de los mismos. Posteriormente, se presenta la implementación de dos versiones del Algoritmo de Selección Negativa (ASN). Antes de utilizar los ASN en la detección de FEII es pertinente realizar un proceso de validación de los mismos. Luego, se ejecutan los ASN en la detección de FEII y los resultados son comparados con los obtenidos al aplicar el enfoque estadístico multivariable.

### **3.1 Simulador UVA/PADOVA**

A continuación, se describen los pasos que permiten replicar los casos de estudio utilizados por Rojas y su equipo (Rojas, R. et al., 2011).

Ya que en el simulador de pacientes diabéticos UVA/PADOVA se cuenta con una variedad de opciones para realizar pruebas, tales como: selección del escenario de pruebas, pacientes objeto de estudio, mediciones de salida, entre otras, se comienza por seleccionar un escenario de tres días con tres comidas regulares cada día. Del grupo de pacientes incluidos en el simulador, se selecciona un grupo de 10 niños y el paciente considerado como paciente promedio, para un total de 11 casos de estudio.

Debido a que se necesita obtener dos conjuntos de datos: un conjunto de datos de pacientes sometidos a terapia continua de infusión de insulina en momentos en los que el equipo de infusión no presente fallas y un segundo conjunto de datos, que corresponde a datos de los mismos pacientes cuando el equipo de infusión falla, en el subsistema de la bomba de insulina del simulador se realizan modificaciones para simular las FEII. Las fallas se generan durante el segundo día del escenario. Se simulan como una degradación en rampa, de 100% a 0%, de la entrega de insulina en un período de 6 horas y sin entrega de bolo de insulina después de la falla. Para obtener la degradación en rampa se hacen modificaciones al archivo *simulink* correspondiente a la plataforma de pruebas del simulador UVA/PADOVA (figura 2.5). Específicamente, en el subsistema correspondiente a la bomba de infusión (pump) se agrega una señal tipo rampa, la cual se multiplica por la salida del módulo que calcula la dosis de insulina justo antes de que esta señal pase a la bomba de inyección. De esta manera, luego de que el algoritmo de control calcule la dosis de insulina, ocurre degradación de la misma, no aplicándose la dosis adecuada al subsistema que representa al paciente (Human Model).

La degradación en rampa de la infusión de insulina se genera a medianoche (12:00 a.m.) cuando no se presenta ninguna comida, al mediodía (12:00 p.m.) simultáneamente con el almuerzo y a las 4:00 p.m., dos horas antes de la cena. Se realiza una simulación por falla, es decir para cada paciente se realizan tres simulaciones, una correspondiente a la falla de medianoche, una segunda simulación, correspondiente a la falla del mediodía y, finalmente, la simulación correspondiente a la falla de la tarde. Adicionalmente, se realiza una simulación en ausencia de fallas para cada paciente.

Luego de tener definidos el escenario de prueba, el grupo de pacientes objeto de estudio y las fallas, se generan los datos. De las estructuras de datos del simulador, se extraen vectores con valores del monitoreo continuo de glucosa, valores correspondientes a la dosis de inyección de insulina, así como también el vector de tiempos de simulación. Se considera un período de muestreo de 5 minutos para simular el muestreo real de los monitores continuos de glucosa, CGM.

### **3.2 Implementación del ASN**

En esta sección se describen los pasos a seguir para la implementación y validación de las dos versiones originales del algoritmo de selección negativa y el procedimiento para selección de parámetros necesarios para la ejecución de los mismos.

#### **3.2.1 ASN con radio de detección fijo**

Este algoritmo es implementado en MATLAB® basado en lo descrito en el marco teórico. Inicialmente, los datos utilizados son normalizados en el intervalo [0,1]. Para este

propósito se programa una rutina en la que se emplea el comando *norm* de MATLAB® aplicando la norma infinita. A continuación, se define el conjunto propio a partir del conjunto de datos que presentan características similares. Los datos que conforman el conjunto propio son seleccionados de forma aleatoria utilizando el comando *randsample* de MATLAB® que permite obtener  $k$ -valores aleatorios muestreados uniformemente y sin reemplazo a partir del conjunto dado. Se genera aleatoriamente el conjunto de tamaño prefijado de "posibles detectores", este proceso se realiza utilizando el comando *rand* de MATLAB® que devuelve una matriz de la dimensión requerida contentiva de valores aleatorios de una distribución uniforme estándar en el intervalo establecido, en este caso [0,1]. Luego, se procede a determinar si cada detector generado coincide con algún dato del conjunto propio. Utilizando estructuras de repetición y decisión se calcula, la distancia Euclíadiana entre cada posible detector y cada dato del conjunto propio. Si la distancia es menor al radio de detección predefinido, el detector es descartado y se genera aleatoriamente uno nuevo. Este bucle finaliza cuando se alcanza el número de detectores predefinido, el cual es obtenido a partir de la ecuación 2.2. Si la distancia es mayor al radio de detección, el detector es almacenado para la siguiente fase del algoritmo (fase de detección). Una vez generado el conjunto de detectores, utilizando estructuras de repetición, cada dato del universo de datos objeto de estudio es comparado con cada detector y, con base en la distancia Euclíadiana, se clasifica como propio o no propio. Si la distancia entre el dato y el detector es menor al radio de detección, el dato es clasificado como no propio, si la distancia es mayor, se clasifica como propio. En cada iteración se va llenando el arreglo de clasificación. Si el  $i$ -ésimo dato es clasificado no propio, se establece "cero" en la posición  $i$  del arreglo y, si el dato es

clasificado como propio, entonces, en esa posición se fija el valor “uno”. Este arreglo es utilizado para el cálculo de tasas de detección y falsos positivos. La tasa de detección es la suma de los puntos no propios detectados. La tasa de falsos positivos es la cantidad de puntos propios detectados incorrectamente durante la ejecución.

### 3.2.2 ASN con radio de detección variable

En el caso del ASN con radio de detección variable, se implementa un pseudocódigo generado a partir de la descripción del algoritmo y luego, es codificado en MATLAB®. De igual manera que en el ASN de radio fijo, los datos utilizados son normalizados en el intervalo [0,1] utilizando el comando *norm* de MATLAB® y aplicando la norma infinita. Haciendo uso del conjunto definido como propio, un porcentaje de datos es elegido aleatoriamente utilizando también el comando *randsample*. Los detectores se generan aleatoriamente, uno a la vez, utilizando el comando *rand* de MATLAB®. Luego, se comprueba que se mantienen a la distancia mínima permitida del conjunto propio. Si la distancia es menor, entonces el detector es eliminado y se genera uno nuevo. De lo contrario, el detector se almacena junto a la distancia calculada, la cual representa su radio variable. A su vez, se incrementa un contador de detectores. Haciendo uso de estructuras de repetición y decisión, de la fórmula de distancia Euclíadiana (ecuación 2.1) y tomando en cuenta el porcentaje de solapamiento permitido, se comprueba que el nuevo detector se encuentra también a una distancia mínima permitida de otros detectores (se utiliza como umbral el radio variable del detector previo almacenado). Si la distancia es menor a la distancia mínima, se genera un nuevo detector. Si no, se incrementa un contador que indica que el

nuevo detector no ha sido detectado por otro y se almacena para la etapa de detección. En cada iteración se va incrementando el contador de cobertura ( $m$ ). Cuando el valor de este contador  $m_{max}$  es alcanzado, el cual es obtenido utilizando la ecuación 2.6 y estableciendo un porcentaje de cobertura estimada deseado ( $c_o$ ), culmina la etapa de generación de detectores.

La fase de detección es similar a la del ASN con detectores de radio fijo, pero en este caso el umbral o mínima distancia permitida hasta cada punto del conjunto propio, se basa en el radio variable asignado a cada detector. Cada dato desconocido cuya distancia hasta el conjunto de detectores sea menor al radio de detección, se clasifica como no propio. En cada iteración se va asignando uno o cero en la posición actual del arreglo de salida dependiendo de si el dato es clasificado como propio o no propio, respectivamente.

### 3.2.3 Validación del ASN

Para validar el funcionamiento de estos algoritmos, se utilizan los resultados presentados por Zhou, Ji. et al., 2004 donde implementan las dos versiones del ASN. Zhou, Ji. et al., ejecutan los ASN con el conjunto de datos Fisher's Iris Dataset (Statlib datasets archive, 2009) ampliamente utilizado en análisis que implican clasificación. El conjunto de datos incluye 150 muestras de tres tipos de flores (50 muestras por cada especie): Iris Setosa, Iris Virginica, Iris Versicolor. Para cada especie de flor se utilizan medidas de cuatro características, largo y ancho del pétalo y del sépalo. Los investigadores se plantean estudiar las propiedades y posibles ventajas que ofrece el ASN con detectores de radio variable. En la fase de generación de ambas versiones del ASN utilizan los datos de una de las especies de flores como conjunto propio, mientras los otros dos son considerados como conjunto no

propio. Los datos propios son utilizados completa y parcialmente en la fase de generación de detectores. El uso de una parte del conjunto propio permite demostrar la capacidad de los algoritmos de reconocer datos propios desconocidos. Los resultados presentados son obtenidos a partir de 100 pruebas repetidas para cada método. El radio de detección utilizado es 0.1, un 99% de cobertura estimada y el máximo de detectores generados es 1000. Para validar los ASN implementados en este trabajo de grado, se utilizan los mismos valores de parámetros. La media y desviación estándar de las tasas de detección, tasas de falsos positivos y número de detectores se muestran en la tabla correspondiente para su análisis y poder determinar si la implementación presentada es válida. El criterio utilizado para validar los ASN consiste en calcular el valor medio porcentual de la diferencia entre las tasas obtenidas y las publicadas y comprobar que es inferior al 10%.

### **3.2.4 Ajuste de parámetros**

#### **3.2.4.1 ASN con radio de detección fijo**

Los principales parámetros del ASN con detectores de radio fijo son el radio de detección y el número máximo de detectores. Ambos son parámetros de diseño que el investigador puede fijar según sus criterios. En nuestro caso, el radio de detección se obtendrá a partir de la distancia Euclídea máxima entre centroide del conjunto propio y cada elemento de este conjunto. Esta distancia máxima, es el punto de partida para el análisis de sensibilidad que permita seleccionar el radio de detección adecuado. El número máximo de detectores se calcula a partir de la ecuación 2.2 y del procedimiento descrito en la sección 2.4.2.

El análisis de sensibilidad para la selección del radio de detección se aplica sobre el conjunto "Fisher's Iris Dataset". Esta fase determina el efecto de la variación de los valores del radio de detección en el desempeño del ASN. Se selecciona el conjunto de datos flores Virginica como conjunto propio. Se escoge como radio de detección base Rdb, la distancia máxima obtenida entre el conjunto de datos propios y el centroide. Luego, se ejecuta el algoritmo variando el radio de detección (se considera el 75%, 100%, 150% y 200% del valor de Rdb). Finalmente, se determina el radio de detección a utilizar, basados en las mejores tasas de detección y falsos positivos.

### **3.2.4.2 ASN con radio de detección variable**

En la versión del ASN con detectores de radio variable se utilizan los siguientes parámetros de diseño: número máximo de detectores, radio de detección, cobertura máxima estimada y porcentaje de solapamiento permitido. Igual que en el caso del ASN de radio fijo, estos parámetros pueden ser fijados por el investigador. En nuestro caso, el radio se obtiene determinando el centroide del conjunto de datos propios y calculando la distancia Euclíadiana que existe entre cada dato de ese conjunto y el centroide. A partir de este procedimiento se obtiene el valor de radio base, Rdb, sobre el cual se realiza el análisis de sensibilidad. El número máximo de detectores se calcula utilizando la ecuación 2.2 y del procedimiento descrito en la sección 2.4.2. Para establecer el valor del parámetro adicional requerido en esta versión, se decide utilizar una cobertura estimada de 99%, ya que se requiere cubrir la mayor parte del espacio no propio. A partir de la ecuación (2.6) se calcula el valor del parámetro  $m$

relacionado con la cobertura estimada y el cual, junto al número máximo de detectores es utilizado como criterio de parada del algoritmo.

Utilizando el conjunto “Fisher’s Iris Dataset” se determina el efecto de la variación de los valores del Rdb en el desempeño del ASN. Se considera el conjunto datos flores Virginica como conjunto propio. Se ejecuta el algoritmo variando el valor del Rdb, se utiliza 75%, 150%, 200% y 250% de este valor. Luego de seleccionar el radio que permita obtener mayores tasas de detección y menores tasas de falsos positivos, se varía el porcentaje de solapamiento permitido entre detectores (25%, 15% y 5%) para estudiar su efecto en el desempeño del algoritmo. Al concluir la etapa de pruebas, se seleccionan los valores definitivos del radio de detección y el porcentaje de solapamiento.

### **3.2.4.3 Desempeño del ASN utilizando nuevos parámetros**

Luego de validar los ASN y fijar nuevos valores para los parámetros, obtenidos a partir de los procedimientos descritos en la sección anterior, se realizan pruebas considerando los datos de cada una de las especies de flores como conjunto propio y los conjuntos de las dos especies restantes como conjunto no propio. Con la finalidad de obtener un mejor desempeño, la inicialización de los parámetros es independiente para cada conjunto de datos. Se realizan 100 pruebas consecutivas en cada caso. La media y desviación estándar de las tasas de falsos positivos y detección se muestran en tablas correspondientes. La evaluación del desempeño de los ASN se realiza con base en los valores de parámetros que permitan obtener un balance entre altas tasas de detección y bajas tasas de falsos positivos.

### 3.2.5 ASN en la detección de FEII

Para la detección de Fallas en los equipos de Infusión de Insulina, FEII, utilizados por pacientes que padecen diabetes tipo I, mediante ASN debemos acondicionar el algoritmo que mostró el mejor desempeño en las pruebas de clasificación de datos como detector de fallas. En este caso la fuente de datos corresponde al monitoreo de glucosa del grupo de pacientes *in silico*, en condiciones normales y en momentos en los que se simulan FEII, generados en el simulador UVA/PADOVA.

Es necesario establecer una diferencia entre la aplicación del ASN en la clasificación de datos (primera etapa de pruebas con el conjunto de flores Iris) y la aplicación del ASN como detector de FEII. La detección de FEII implica el manejo de datos de serie de tiempo. La estructura de datos de los pacientes simulados incluyen el vector de datos del monitoreo continuo de glucosa y el vector de tiempos de simulación. Los datos de interés para nuestra investigación, generados en el simulador, tienen la estructura de un vector columna contentivo de valores correspondientes al monitoreo de glucosa en plasma en períodos de simulación predefinidos. Se considera un período de muestreo de 5 minutos para simular el muestreo real de los monitores continuos de glucosa. En este contexto, para la ejecución del ASN es necesario considerar vectores que contengan los últimos 45 minutos de datos para glucosa en plasma y tiempo (10 datos).

Los datos requieren ser normalizados antes de ser utilizados en el algoritmo. La rutina de normalización utilizada para este propósito es la misma utilizada en la etapa de pruebas

con el conjunto de flores Iris, realizando las adaptaciones necesarias debido a que la dimensión del conjunto es distinta. Se selecciona el valor del radio de detección base, Rdb, obteniendo el centroide del conjunto de datos normales (conjunto propio) y calculando la distancia desde cada dato del conjunto hasta el centroide. Se elige como radio de detección base (Rdb) la distancia mínima calculada. Posteriormente, se realiza un análisis de sensibilidad, utilizando los datos correspondientes al paciente promedio, variando el Rdb (se utilizan el 50%, 75% 100% y 150% de este valor). El radio de detección se elige tomando en cuenta las menores tasas de falsos positivos y mayores tasas de detección. El ASN, dependiendo de la versión que se implemente, requiere que se asigne valores a los parámetros: número máximo de detectores, cobertura estimada, porcentaje de solapamiento permitido entre detectores y radio de detección. La selección de estos valores se realiza replicando el procedimiento utilizado el conjunto de datos de flores Iris. En caso de aplicar la versión de radio variable, se establece una cobertura estimada de 99.99% ya que se requiere cubrir la mayor parte del espacio no propio, a partir de este valor, se calcula el parámetro  $m$  (utilizado como criterio de parada del algoritmo). El valor del porcentaje de solapamiento permitido seleccionado es el que mejores tasas de detección y menores tasas de falsos positivos genera en el análisis de sensibilidad, en el cual se varía este valor (25%, 15% y 5%). El número máximo de detectores se calcula a partir de la ecuación 2.2. Una vez establecidos los parámetros a utilizar, se procede a ejecutar el ASN.

La primera fase del ASN, etapa de generación de detectores, es una etapa en la que un porcentaje del conjunto total de datos normales es seleccionado, aleatoriamente, para formar

el conjunto propio. Se utiliza el 75% de los datos en condiciones normales. Para la selección aleatoria, hacemos uso de la rutina implementada en MATLAB® utilizada en las pruebas anteriores, nuevamente tomando en consideración que las dimensiones de los conjuntos de datos son diferentes. Cada dato de este conjunto propio es comparado con el conjunto de detectores generado aleatoriamente. Esta comparación consiste en calcular la distancia Euclidiana entre cada punto propio y cada detector y si la misma es menor que la distancia mínima permitida, el detector es descartado y se genera uno nuevo, de lo contrario el detector es almacenado. En el caso de ejecutarse la versión de radio variable, el detector se almacena junto a la distancia calculada, que se asigna como su radio variable. Se incrementa el contador de detectores. Tomando en cuenta el porcentaje de solapamiento, se comprueba que este nuevo detector se encuentra también separado de otros detectores. Si la distancia es menor a la distancia mínima, se genera un nuevo detector. Si no, éste se almacena para la etapa de detección. En cada iteración se va incrementando el contador de cobertura ( $m$ ). Cuando el valor de este contador  $m_{max}$  es alcanzado, culmina la fase de generación.

En la fase de detección se consideran los datos muestreados a partir del monitoreo continuo de glucosa en momentos en los que se simulan las FEII. Se selecciona una banda de valores de glucosa en plasma entre cero y 500 mg/dL. Cada valor muestreado es comparado con cada dato del conjunto de detectores generado en la etapa anterior. La comparación se realiza calculando la distancia Euclidiana entre cada dato y si esta es menor al radio de detección definido, el dato es clasificado como no propio, de lo contrario es clasificado como propio. En cada iteración se va almacenando “uno” o “cero” en la posición actual del arreglo

de salida según el dato sea clasificado como propio o no propio, respectivamente. Este arreglo se utiliza para el cálculo de tasas de detección y falsos positivos. Además, se genera un arreglo contentivo de los datos clasificados como no propios y un arreglo de datos clasificados como propios. Cada dato debe pasar por el proceso inverso a la normalización inicial para hallar su equivalente en valor de glucosa en plasma y posteriormente ser utilizado en la generación de las gráficas correspondientes. En este contexto, clasificar el dato como propio o no propio, es equivalente a detectar un punto normal o un punto de falla, respectivamente.

### **3.3 Comparación entre ASN y Enfoque Estadístico Multivariable**

La comparación entre los resultados obtenidos aplicando el ASN de radio variable y los de la literatura (Rojas, García y Bequette, 2011) obtenidos aplicando un enfoque estadístico multivariable se realiza con base en la tasa de falsos positivos por hora y tiempo de detección. Para esto se organizan dos tablas resumen, donde se muestran los resultados de esta investigación y los publicados.

# **Capítulo 4**

## **Análisis de resultados**

En este capítulo se presentan los resultados obtenidos al manejar el simulador UVA/PADOVA para generar el conjunto de datos utilizados. De igual manera, se presentan los resultados obtenidos al implementar y validar el ASN, por último, se presentan los resultados obtenidos al utilizar el ASN como método detector de FEII.

### **4.1 Simulador UVA/PADOVA**

En figura 4.1, se observa la gráfica correspondiente a los valores de insulina para el paciente promedio simulado. En negro se aprecia la curva que describe un comportamiento normal de los niveles de insulina. Nótese que los momentos en los que ocurre una elevación de los valores de insulina, son justamente en las horas donde el paciente consume alimentos (desayuno, almuerzo y cena), lo que ocurre con la finalidad de mantener los valores de glucosa en plasma dentro de los valores normales, los cuales se alteran con la ingesta alimenticia. En color verde se observa la degradación en rampa que ocurre a la dosis de insulina en el momento en el que se genera la falla de medianoche, los valores de insulina comienzan a degradarse con el transcurso. En rojo se presenta la degradación que ocurre en la dosis de insulina cuando se genera la falla del mediodía y en color azul la degradación de la dosis que ocurre al generar la falla de las cuatro de la tarde.

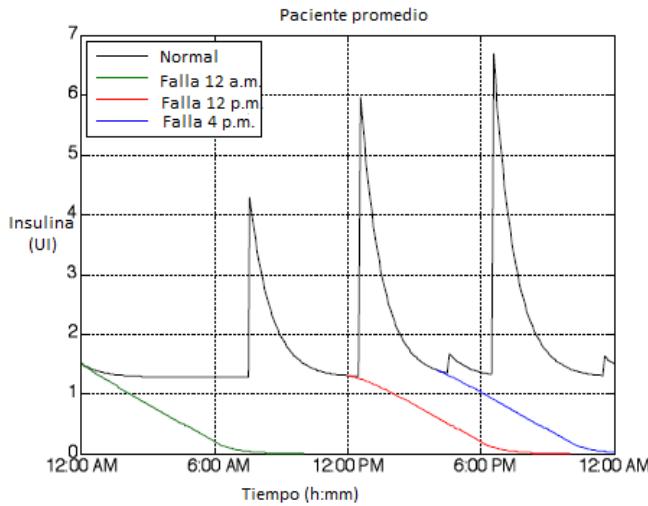


Figura 4. 1 Degradación en rampa de la dosis de insulina para el paciente promedio simulado

A continuación, se presenta la gráfica correspondiente a los datos obtenidos al realizar las simulaciones del paciente #4. Se grafica, simultáneamente, la señal del sensor de monitoreo continuo de glucosa en momentos en los que no estaban ocurriendo FEII y la señal en momentos en los que se presentan las fallas. En la gráfica color negro se aprecian los valores de glucosa en plasma en ausencia de FEII. En color verde, rojo y azul, los valores de glucosa en plasma después de producirse las fallas de medianoche, mediodía y 4:00 de la tarde, respectivamente.

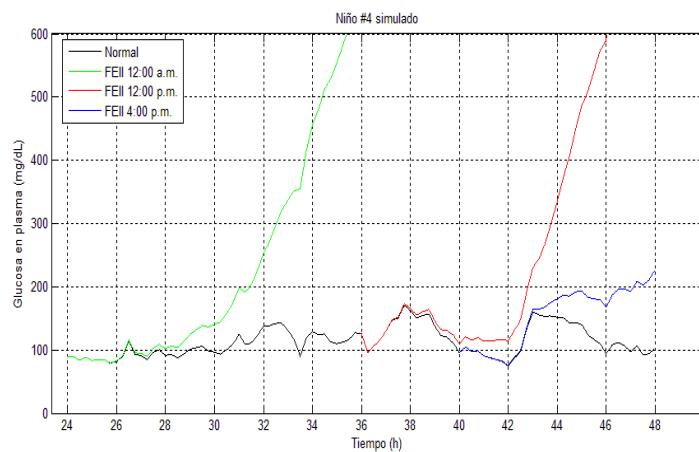


Figura 4. 2 Respuesta de la glucosa en plasma para las tres fallas simuladas. Paciente simulado #4.

En la figura 4.3 se presenta el módulo de la bomba de insulina donde se incluye la rampa de degradación de la infusión de insulina. Como se puede observar, se agrega al módulo una señal tipo rampa (observar el recuadro rojo) la cual es multiplicada por la señal de insulina total (dosis calculada) y el producto obtenido corresponde a la señal que recibe la bomba de infusión de insulina. La bomba de insulina recibe, por lo tanto, una señal que va siendo atenuada al transcurrir el tiempo.

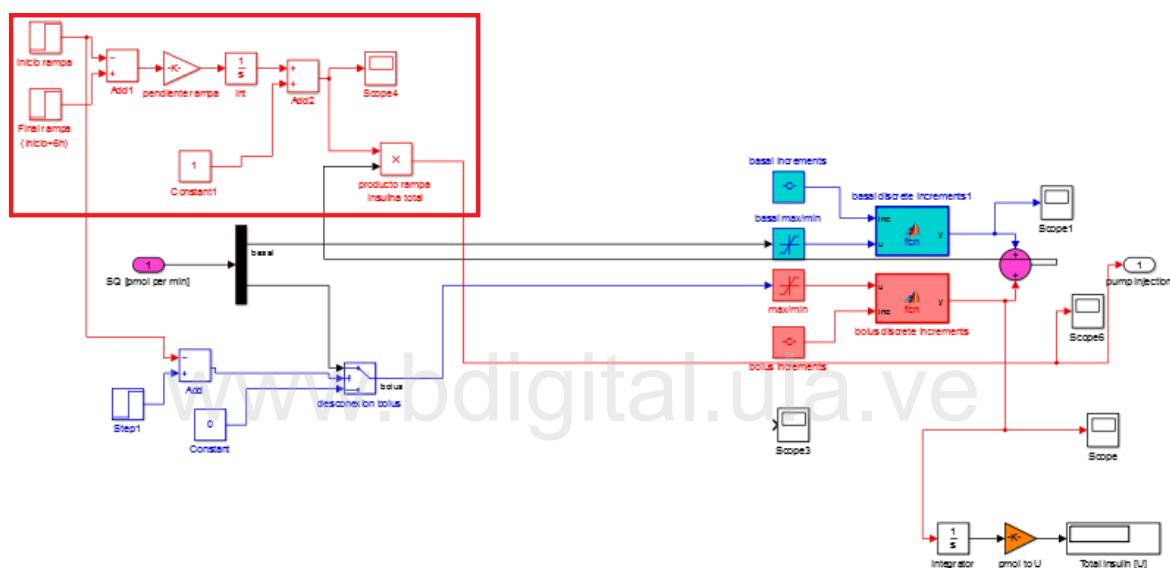


Figura 4. 3 Módulo correspondiente a la bomba de infusión de insulina del simulador UVA/PADOVA incluyendo la rampa de degradación de insulina.

## 4.2 Implementación del ASN

#### **4.2.1 ASN con radio de detección fijo**

El pseudocódigo obtenido para el ASN con detectores de radio fijo es el siguiente.

## Cargar conjunto total de datos

### Cargar conjunto propio

#### Selección tamaño conjunto propio

Ejecutar rutina de normalización

Ejecutar rutina de normalización  
Ejecutar rutina de selección aleatoria

---

Parámetros predefinidos:  $r_s$ ,  $D_{max}$

Inicio

Mientras ( $i < D_{max}$ )

- Generar aleatoriamente el conjunto de detectores,  $d_i$
- Para cada muestra propia
  - Calcular la distancia a cada detector, dist
  - Si (dist  $< r_s$ )
    - Generar aleatoriamente un nuevo de detector,  $d_i$
  - Si no
    - Almacenar detector como  $d_i$
    - Incrementar i+1
- Fin Si

Fin Para

Fin Mientras

Fin

Para cada dato desconocido  $d_j$

Hacer dist\_min igual a infinito

Para cada punto del conjunto de detectores,  $d_i$

- Calcular la distancia (dist) a cada dato desconocido
- Si (dist  $<$  dist\_min)
  - Almacenar dist como la nueva dist\_min
  - Si (dist\_min  $< r_j$ )
    - Almacenar 0 en posición j del vector *conjunto\_propio*
  - Si no
    - Almacenar 1 en posición j del vector *conjunto\_propio*

Fin Si

Fin Si

Fin Para

Fin Para

Fin

#### 4.2.2 ASN con radio de detección variable

El pseudocódigo utilizado para implementar el ASN con detectores de radio variable es el siguiente:

Cargar conjunto de datos propios  
 Selección tamaño conjunto propio  
 Ejecutar rutina de normalización  
 Ejecutar rutina de selección aleatoria

Parámetros de control predefinidos:  $r_s$ ,  $m_{max}$ ,  $D_{max}$

Inicio

Mientras ( $m < m_{max}$ ) || ( $i < D_{max}$ )

- Generar aleatoriamente el conjunto de detectores,  $d_i$
- Calcular la distancia más corta a cada punto propio, dist\_min
- Si ( $dist\_min < r_s$ )
  - Retornar al inicio
- Si no
  - Si ( $i=1$ )
    - Almacenar detector como  $d_i$  y  $r_{di} = dist\_min$
    - Incrementar  $i+1$
  - Si no
    - Calcular la distancia mínima para cada detector anterior, dist\_min2
    - Si ( $dist\_min2 < r_{di}$ )
      - $m = m+1$
    - Si no
      - Almacenar detector como  $d_i$  y  $r_{di} = dist\_min$
      - Incrementar  $i+1$
      - $m = 0$

- Fin Si
- Fin Si
- Fin Mientras
- Fin

Para cada dato desconocido  $d_j$

Hacer dist\_min igual a infinito

Para cada punto del conjunto de detectores,  $d_i$

- Calcular la distancia (dist) a cada dato desconocido
- Si ( $dist < dist\_min$ )
  - Almacenar dist como la nueva dist\_min
  - Si ( $dist\_min < r_{di}$ )
    - Almacenar 0 en posición  $j$  del vector *conjunto\_propio*
- Si no
  - Almacenar 1 en posición  $j$  del vector *conjunto\_propio*

- Fin Si
- Fin Si
- Fin Para
- Fin Para
- Fin

#### 4.2.3 Validación del ASN

A continuación, se presenta la distribución de los datos correspondientes al conjunto de datos Fisher's Iris (figuras 4.4 y 4.5). Inicialmente, se grafican las dos primeras dos características de cada tipo de flor (largo y ancho del sépalo) y, posteriormente, se grafican la tercera y cuarta característica de cada tipo de flor (largo y ancho del pétalo). Como se puede apreciar en las gráficas, los datos de la flor Iris Setosa están claramente diferenciados del resto de los datos, hecho que no se aprecia en el caso de las otras dos especies donde la distribución es menos definida.

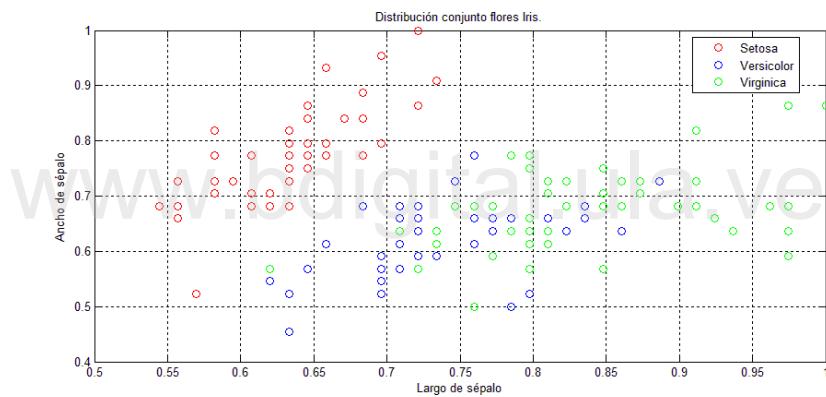


Figura 4. 4 Distribución del ancho y largo del sépalo. Conjunto de datos Iris.

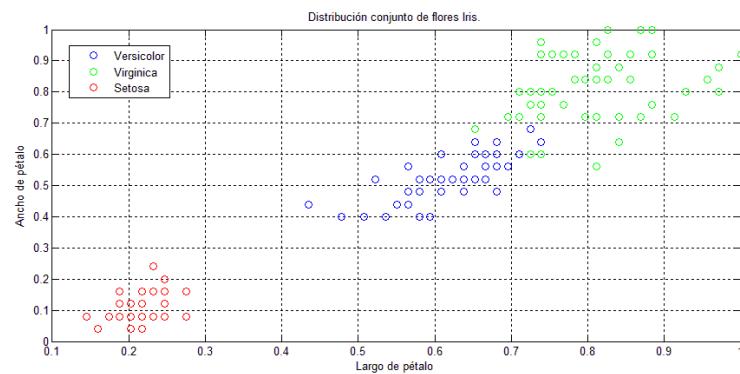


Figura 4. 5 Distribución del ancho y largo del pétalo. Conjunto de datos Iris.

En la tabla 4.1 se muestran los resultados obtenidos por Zhou, Ji et al., y los resultados obtenidos en esta investigación al aplicar el ASN con detectores de radio fijo utilizando los mismos valores de los parámetros (radio de detección: 0.1 y número máximo de detectores: 1000). Como se puede apreciar, en ambas implementaciones, las mayores tasas de detección se obtienen al utilizar una parte del conjunto propio en la fase de generación (25 datos). El error absoluto calculado entre las tasas publicadas y las obtenidas es inferior al 2.5% en todos los casos, llegando incluso a valores por debajo del 0.2%.

Tabla 4. 1 Tasa de detección obtenida al aplicar el ASN de radio fijo utilizando conjunto de datos flores Iris.

Conjunto de Datos	Número de datos	Tasa de detección		Tasa de detección		Error absoluto %	
	Propios	(Zhou, Ji 2004) (%)		Obtenida (%)			
		Media	DE	Media	DE		
Setosa	50	100.00	0.00	99.40	1.51	0.60	
Virginica	50	92.51	0.74	94.77	3.88	2.44	
Versicolor	50	95.67	0.69	94.26	2.92	1.47	
Setosa	25	100.00	0.00	99.55	0.90	0.45	
Virginica	25	97.18	0.71	96.87	2.83	0.32	
Versicolor	25	96.00	0.45	96.18	3.01	0.19	

Tabla 4. 2 Tasa de falsos positivos obtenidas al ejecutar el ASN de radio fijo, utilizando conjunto de datos flores Iris.

Conjunto de Datos	Número de datos	Tasa de falsos positivos		Tasa de falsos positivos obtenida		Error absoluto %	
	Propios	(Zhou, Ji 2004)		obtenida			
		Media	DE	Media	DE		
Setosa	50	0.00	0.00	0.00	0.00	0.00	
Virginica	50	0.00	0.00	0.00	0.00	0.00	
Versicolor	50	0.00	0.00	0.00	0.00	0.00	
Setosa	25	11.18	2.17	10.78	6.48	3.58	
Virginica	25	33.26	0.96	34.48	5.77	3.67	
Versicolor	25	22.20	1.25	22.40	7.84	0.90	

Los resultados que se presentan en la tabla 4.2 corresponden a las tasas de falsos positivos obtenidas al ejecutar el ASN con detectores de radio fijo. Igual que en el caso anterior, los valores de los parámetros utilizados son: radio de detección: 0.1, número total de detectores generados: 1000. Las tasas de falsos positivos son mayores cuando se utilizan 25 datos como conjunto propio. El error absoluto porcentual obtenido entre los valores publicados y los obtenidos es inferior al 5% en todos los casos.

Tabla 4. 3 Tasa de detección obtenida al ejecutar el ASN de radio variable. Utilizando conjunto de datos flores Iris.

Conjunto de Datos	Número de datos Propios	Tasa de detección (Zhou, Ji 2004)		Tasa de detección obtenida		Error absoluto %
	Media	DE	Media	DE		
Setosa	50	99.98	0.14	99.92	0.40	0.06
Virginica	50	81.87	2.78	89.19	5.52	8.94
Versicolor	50	85.95	2.44	84.16	5.49	2.08
Setosa	25	99.97	0.17	99.96	0.31	0.01
Virginica	25	93.58	2.33	95.53	4.59	2.08
Versicolor	25	88.30	2.77	83.49	4.79	5.45

Los resultados que se presentan en la tabla 4.3 corresponden a las tasas de detección obtenidas al ejecutar el ASN con detectores de radio variable. Se muestran los resultados obtenidos por Zhou, Ji et al., y los obtenidos en esta investigación. Los valores de los parámetros utilizados son: radio de detección: 0.1, número total de detectores generados: 1000, cobertura estimada: 99% y porcentaje de solapamiento: 5%. El porcentaje de solapamiento permitido no es un dato aportado en la publicación, por tal razón fue necesario

ajustar el valor de este parámetro para obtener tasas cercanas a las publicadas. Utilizando el conjunto de flores Virginica y 25 datos como conjunto propio, el error absoluto porcentual es el más elevado (superior al 8%), en los otros casos se mantiene inferior al 5%. Los valores del error absoluto porcentual más elevados en esta versión del algoritmo son atribuibles al desconocimiento del valor real del porcentaje de solapamiento utilizado por los investigadores.

Tabla 4. 4 Tasa de falsos positivos obtenidas al ejecutar el ASN de radio variable. Utilizando conjunto de datos flores Iris.

Conjunto de Datos	Número de datos propios	Tasa de falsos positivos (Zhou, Ji 2004)		Tasa de falsos positivos obtenida		Error absoluto %
		Media	DE	Media	DE	
Setosa	50	0.00	0.00	0.00	0.00	0.00
Virginica	50	0.00	0.00	0.00	0.00	0.00
Versicolor	50	0.00	0.00	0.00	0.00	0.00
Setosa	25	1.32	0.95	2.14	0.70	6.21
Virginica	25	13.18	3.24	12.16	5.67	7.74
Versicolor	25	8.42	2.12	7.40	3.65	12.11

Los resultados que se presentan en la tabla 4.4 corresponden a las tasas de falsos positivos obtenidas al ejecutar el ASN con detectores de radio variable. Los valores de los parámetros utilizados son los mismos que en el caso anterior. En la versión del ASN con detectores de radio variable ejecutada en esta investigación, la tasa de falsos positivos fue cero en todos los casos en los que se utilizó el total de datos del conjunto propio (50 datos), igual que en el estudio publicado. Las tasas de falsos positivos son mayores cuando se utilizan 25 datos como conjunto propio. En el caso del conjunto de flores Versicolor y

utilizando 25 datos como conjunto propio, el error absoluto porcentual es el más elevado (superior al 12%), en los otros casos se mantiene inferior al 8%.

Tabla 4. 5 Número de detectores generados al ejecutar el ASN de radio variable. Utilizando conjunto de datos flores Iris.

Conjunto de Datos	Número de datos propios	Número de detectores		Número de detectores obtenidos		Error absoluto %
		Media	DE	Media	DE	
Setosa	50	20.00	7.87	16.34	5.43	18.30
Virginica	50	218.36	66.11	261.09	37.35	19.56
Versicolor	50	153.24	38.8	161.19	33.11	5.19
Setosa	25	16.44	5.63	14.00	2.09	14.84
Virginica	25	108.12	30.74	112.72	22.65	4.25
Versicolor	25	110.08	22.61	118.21	33.24	7.38

En la tabla 4.5 se resumen los resultados obtenidos al aplicar el ASN con detectores de radio variable, los resultados corresponden a la cantidad de detectores generados. Se muestran los resultados obtenidos en la publicación de Zhou, Ji et al., y los obtenidos en muestra investigación. El error absoluto porcentual entre la cantidad de detectores generados alcanza un valor de casi 20%. La cantidad de detectores generados tiene relación directa con el porcentaje de solapamiento que se permita, como este dato no es aportado por los investigadores en su publicación, el error absoluto porcentual puede atribuirse al ajuste que fue necesario hacer para determinar el valor de este parámetro.

Analizando las tablas correspondientes a las tasas de detección y falsos positivos (tablas 4.1 a 4.4), es importante mencionar que el promedio de error absoluto porcentual para las tasas de detección utilizando 50 datos es 2.60%, utilizando 25 datos como conjunto

propio, este promedio es 2.01%. En el caso de las tasas de falsos positivos, el promedio de error absoluto para los casos en los que se utilizó 50 datos es 0%. Cuando se utilizan 25 datos el promedio de error es 5.71%. Finalmente, se puede concluir que los algoritmos implementados en esta investigación son válidos, ya que cumplen con el criterio propuesto en el cual se acepta un error absoluto porcentual, entre tasas de detección y falsos positivos publicadas y obtenidas, menor al 10%.

#### **4.2.4 Ajuste de parámetros**

##### **4.2.4.1 ASN con detectores de radio fijo**

A continuación se presentan los resultados obtenidos al realizar el procedimiento para el ajuste de parámetros. Se aplica el ASN de radio fijo haciendo variaciones en el radio de detección base, Rdb, y utilizando el conjunto de flores Virginica como conjunto propio.

Tabla 4. 6 Resultados obtenidos al ejecutar el ASN con detectores de radio fijo utilizando el conjunto Iris Virginica como conjunto propio. Variación del radio base de detección. Tasas de detección.

<i>Conjunto de Datos</i>	<i>Número de datos Propios</i>	<i>Tasa de Detección Media</i>	<i>% Radio de detección base</i>
	50	98.78	0.70
Iris Virginica	50	98.05	1.00
	50	96.12	2.33
	50	94.84	4.92
	25	99.62	0.56
	25	99.05	0.96
	25	97.73	1.89
	25	97.44	3.08

En la tabla 4.6 se muestran los valores de media y desviación estándar de las tasas de detección al variar el radio de detección base. Las tasas de detección se mantienen en todas las pruebas por encima del 94% y son mayores mientras menos datos se usen como conjunto propio.

Tabla 4. 7 Resultados obtenidos al ejecutar el ASN con detectores de radio fijo utilizando el conjunto Iris Virginica como conjunto propio. Variación del radio de detección base (Rdb). Tasas de falsos positivos.

<i>Conjunto de Datos</i>	<i>Número de datos</i>	<i>Tasa de falsos positivos</i>		<i>% Radio de detección base</i>
	<i>Propios</i>	<i>Media</i>	<i>DE</i>	
Iris Virginica	50	8.06	11.78	75
	50	13.70	6.02	100
	50	9.12	18.84	150
	50	15.92	25.76	200
	25	34.04	13.63	75
	25	24.14	15.50	100
	25	26.14	19.80	150
	25	16.38	8.64	200

En la tabla 4.7 se muestra los resultados correspondientes a la media y desviación estándar de la tasa de falsos positivos obtenidos al aplicar el ASN de radio fijo, variando el radio de detección base. El número de datos utilizados como conjunto propio y las variaciones del radio base, causan que las tasas de falsos positivos varíen desde un valor superior al 34%, cuando se utiliza el 75% del radio de detección base y se utilizan 25 datos como conjunto propio, hasta un mínimo de 8.06% cuando se utiliza el 75 % del Rdb y se usa el total de datos como conjunto propio (50 datos). Las tasas de falsos positivos son inferiores mientras más datos se usen como conjunto propio.

#### 4.2.4.2 ASN con detectores de radio variable

A continuación se presentan los resultados correspondientes a media y desviación estándar de las tasas de detección y de falsos positivos, respectivamente, obtenidas al aplicar el ASN de radio variable utilizando el conjunto de flores Virginica como conjunto propio, cambiando el radio de detección base y manteniendo el porcentaje de solapamiento en 25%.

Tabla 4. 8 Resultados obtenidos al ejecutar el ASN con detectores de radio variable utilizando el conjunto Iris Virginica como conjunto propio. Variación del Rdb. Tasas de detección.

<i>Conjunto de Datos</i>	<i>Número de datos</i>	<i>Tasa de Detección</i>		<i>% Radio de detección base</i>
	<i>Propios</i>	<i>Media</i>	<i>DE</i>	
Iris Virginica	50	98.57	0.62	75
	50	99.07	0.77	100
	50	97.17	1.16	150
	50	95.80	2.17	200
	25	99.50	0.67	75
	25	98.00	0.63	100
	25	98.30	1.55	150
	25	98.20	1.05	200

Tabla 4. 9 Resultados obtenidos al ejecutar el ASN con detectores de radio variable utilizando el conjunto Iris Virginica como conjunto propio. Variación del Rdb. Tasas de falsos positivos.

<i>Conjunto de Datos</i>	<i>Número de datos</i>	<i>Tasa de Falsos positivos</i>		<i>% Radio de detección base</i>
	<i>Propios</i>	<i>Media</i>	<i>DE</i>	
Iris Virginica	50	0.00	0.00	75
	50	0.00	0.00	100
	50	0.00	0.00	150
	50	0.00	0.00	200
	25	26.93	6.71	75
	25	22.27	6.10	100
	25	15.60	5.30	150
	25	9.40	3.27	200

Analizando los resultados mostrados en las tablas 4.8 y 4.9, las variaciones del radio de detección base y el número de datos utilizados como conjunto propio causan que las tasas de falsos positivos varíen desde un valor medio superior al 26%, cuando se utiliza el 75% del radio base y 25 datos como conjunto propio, hasta una media mínima de 9.40% cuando se utiliza el doble del radio de detección base y se utilizan 25 datos. Las tasas de falsos positivos pueden reducirse al 0% cuando se utiliza el total de datos del conjunto propio en la fase de generación del algoritmo. La tasa de detección se mantiene en todas las pruebas por encima del 95%. Las mayores tasas de detección se obtienen al utilizar menos datos como conjunto propio.

En las tablas 4.10 y 4.11 se muestran la media y desviación estándar de las tasas de detección y falsos positivos, respectivamente, obtenidas al aplicar el ASN con detectores de radio variable. En este caso se utiliza el doble del radio de detección base ( $R_{db}$ ) y se varía el porcentaje de solapamiento permitido para evaluar su efecto en el desempeño del algoritmo. Se utiliza el doble del radio de detección base ya que fue el radio que permitió lograr menores tasas de falsos positivos y mayores tasas de detección. En este sentido, debido a que el 5 % de solapamiento ofrece mejores tasas de detección y menores tasas de falsos positivos, se establece este valor para la siguiente fase de pruebas. La selección del valor de los parámetros obedece a la finalidad con la que se utilice el algoritmo. Si la prioridad es que la detección sea totalmente precisa, es posible lograrlo a costa de una tasa mayor de falsos positivos. Por otro lado, si un alto porcentaje de falsos positivos no es conveniente, es posible lograr un valor adecuado, permitiendo tasas de detección menores.

Tabla 4. 10 Resultados obtenidos ejecutando el ASN con detectores de radio variable utilizando el conjunto Virginica como conjunto propio. Variación del porcentaje de solapamiento. Tasa de detección.

<i>Conjunto de Datos</i>	<i>Número de datos Propios</i>	<i>Tasa de Detección Media</i>	<i>DE</i>	<i>% Radio de solapamiento</i>
Iris Virginica	50	95.80	2.17	25
	50	94.70	2.37	15
	50	93.17	4.37	5
	25	98.20	1.05	25
	25	97.73	1.98	15
	25	96.00	4.30	5

Tabla 4. 11 Resultados obtenidos ejecutando el ASN con detectores de radio variable utilizando el conjunto Iris Virginica como conjunto propio. Variación del porcentaje de solapamiento. Tasa de falsos positivos.

<i>Conjunto de Datos</i>	<i>Número de datos Propios</i>	<i>Tasa de Falsos positivos Media</i>	<i>DE</i>	<i>% Radio de solapamiento</i>
Iris Virginica	50	0.00	0.00	25
	50	0.00	0.00	15
	50	0.00	0.00	5
	25	9.40	3.27	25
	25	10.00	4.87	15
	25	8.47	6.23	5

#### 4.2.4.3 Desempeño del ASN utilizando nuevos parámetros

En esta sección se presentan resultados correspondientes a la ejecución de los ASN utilizando los parámetros seleccionados con base en los análisis realizados en la sección inmediata anterior. Se presentan los resultados correspondientes a los mejores desempeños de ambas versiones el ASN.

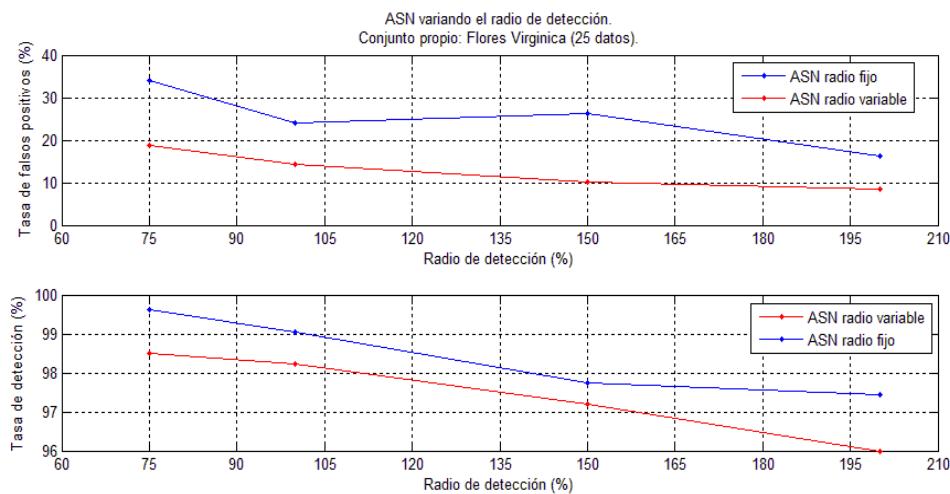


Figura 4.6 . Efectos de la variación del radio de detección sobre las tasas de detección y falsos positivos.

En la figura 4.6 se muestran las gráficas correspondientes a los resultados obtenidos al ejecutar el ASN con detectores de radio fijo y variable cambiando el radio de detección. El conjunto de flores Virginica conforma el conjunto propio y se utiliza la mitad del conjunto (25 datos) en la fase de generación de detectores. Se observa que las tasas de falsos positivos son menores cuando se ejecuta el ASN con detectores de radio variable, pero, al aumentar el radio de detección, las tasas de falsos positivos disminuyen para ambas versiones del algoritmo. Las tasas de detección disminuyen conforme se utiliza un radio de detección mayor.

En la tabla 4.12 se presentan los resultados al ejecutar ambas versiones del ASN utilizando el conjunto de datos de las flores Iris, pero seleccionando los parámetros a partir de la descripción detallada en la sección relativa al ajuste de parámetros (sección 3.2.4).

Como se puede observar, las tasas de detección se mantienen por encima 96%, lo cual representa una mejora con respecto a la ejecución de los algoritmos con valores de parámetros propuestos en la investigación de Zhou, Ji et al., en la que la tasa de detección puede llegar a una media mínima de 81.87%. Las tasas de falsos positivos (tabla 4.13) pueden llegar a 0% cuando se utiliza el total de datos del conjunto propio, cosa que también se observa en los resultados publicados por Zhou, Ji et al. Cuando se usan menos datos como conjunto propio, las tasas de falsos positivos se elevan.

Tabla 4. 12 Resultados obtenidos aplicando ASN de radio fijo y variable. Utilizando conjunto de datos flores Iris y los parámetros calculados. Tasas de detección.

Conjunto de Datos	Número de datos		Tasa de Detección		Tasa de Detección Radio Variable	
			Radio Fijo			
	Propios	Media	DE	Media		
Setosa	50	100.00	0.00	100.00	0.00	
Virginica	50	98.03	1.22	96.60	1.54	
Versicolor	50	98.96	1.02	96.40	1.56	
Setosa	25	100.00	0.00	100.00	0.00	
Virginica	25	98.99	1.03	98.23	1.23	
Versicolor	25	99.51	0.77	98.13	1.63	

Tabla 4. 13 Resultados obtenidos aplicando ASN de radio fijo y variable. Utilizando conjunto de datos flores Iris y los parámetros calculados. Tasas de falsos positivos.

Conjunto de Datos	Número de datos		Tasa de falsos positivos		Tasa de falsos positivos Radio Variable	
			Radio Fijo			
	Propios	Media	DE	Media		
Setosa	50	9.16	3.44	0.00	0.00	
Virginica	50	6.92	1.18	0.00	0.00	
Versicolor	50	9.96	3.35	0.00	0.00	
Setosa	25	18.18	2.13	6.53	3.10	
Virginica	25	22.50	1.58	14.20	7.05	
Versicolor	25	36.94	1.69	21.53	6.30	

En la tabla 4.14 se muestran la cantidad de detectores generados en ambas versiones del ASN utilizando los parámetros calculados en la sección inmediata anterior. La versión del ASN con detectores de radio variable genera una cantidad menor de detectores.

Tabla 4. 14 Resultados obtenidos aplicando ASN de radio fijo y variable. Utilizando conjunto de datos flores Iris y los parámetros calculados. Número de detectores.

Conjunto de Datos	Número de datos Propios	Número de detectores		Número de detectores	
		ASN radio variable		ASN radio fijo	
		Media	DE	Media	DE
Setosa	50	131.64	21.97	1212	166.37
Virginica	50	121.54	22.89	1154	143.66
Versicolor	50	122.52	32.58	1174	180.67
Setosa	25	121.39	21.75	1227	158.47
Virginica	25	121.11	23.13	1233	118.42
Versicolor	25	119.21	32.97	1260	180.00

En este contexto, habiendo determinado que el desempeño del ASN con detectores de radio variable ofrece ventajas en cuanto a menores tasas de falsos positivos y tasas mayores de detección, además de ser una versión más eficiente ya que requiere una menor cantidad de detectores, lo que aumenta la velocidad de ejecución del algoritmo, se decide implementar esta versión del algoritmo en la fase de detección de FEII.

#### 4.2.5 ASN en la detección de FEII

Con base en los resultados obtenidos en la sección anterior, se utiliza el ASN con detectores de radio variable para la detección de FEII.

En la tabla 4.15 se presentan los resultados obtenidos al variar el radio de detección base, aplicando el ASN con los datos correspondientes al paciente promedio cuando se simulan las tres fallas. Se puede observar que las mayores tasas de detección y menores tasas de falsos positivos se obtienen utilizando el 50% del Rbd. Por lo tanto, este valor es seleccionado para la siguiente etapa de pruebas.

Tabla 4. 15 Efectos de la variación del Rbd en las tasas de detección y falsos positivos.

<i>Caso de estudio</i>	<i>Falla 12:00 a.m.</i>		<i>Falla 12:00 p.m.</i>		<i>Falla 12:00 p.m.</i>		<i>Rbd (%)</i>
	<i>Tasa detección</i>	<i>Tasa falsos positivos</i>	<i>Tasa detección</i>	<i>Tasa falsos positivos</i>	<i>Tasa detección</i>	<i>Tasa falsos positivos</i>	
Paciente Promedio	96.43	3.57	100.00	7.14	88.89	22.32	50
	92.86	7.29	92.31	14.29	77.78	25.16	75
	89.29	10.71	92.31	21.43	66.67	26.32	100
	84.88	7.14	92.31	12.94	44.44	21.16	150

En la tabla 4.16 se presentan los resultados al variar el porcentaje de solapamiento utilizando el valor del Rdb calculado en la fase anterior. Se ejecuta el ASN con los datos del paciente promedio en momentos en los que se simulan las tres fallas. Se puede observar que permitiendo el 25 % de solapamiento se obtienen las mayores tasas de detección y menores tasas de falsos positivos.

Tabla 4. 16 Efectos de la variación del porcentaje de solapamiento en las tasas de detección y falsos positivos.

<i>Caso de estudio</i>	<i>Falla 12:00 a.m.</i>		<i>Falla 12:00 p.m.</i>		<i>Falla 12:00 p.m.</i>		<i>Solapamiento (%)</i>
	<i>Tasa detección</i>	<i>Tasa falsos positivos</i>	<i>Tasa detección</i>	<i>Tasa falsos positivos</i>	<i>Tasa detección</i>	<i>Tasa falsos positivos</i>	
Paciente Promedio	96.43	3.57	100.00	7.14	88.89	22.32	25
	82.14	17.86	92.31	14.29	100.00	26.32	15
	78.57	21.43	92.31	28.57	100.00	36.32	5

En la tabla 4.17 se presentan los resultados correspondientes a las tasas de detección obtenidas al aplicar el ASN de radio variable en la detección de las FEII en el grupo de 11 pacientes simulados con los parámetros seleccionados (50% del Rdb, 25% de solapamiento permitido). Los resultados son obtenidos utilizando los datos de cada paciente en el momento en el que se模拟aron tres fallas (a medianoche, a las 12:00 del mediodía y a las 4:00 de la tarde).

Tabla 4. 17 Tasa de detección porcentuales aplicando ASN para cada paciente simulado (%).

<i>Caso de estudio</i>	<i>FEII de medianoche (12:00 a.m.)</i>	<i>FEII de mediodía (12:00 p.m.)</i>	<i>FEII 4 de la tarde (4:00 p.m.)</i>
Paciente #1	98.75	92.31	72.22
Paciente #2	92.86	76.75	81.11
Paciente #3	89.29	92.15	77.78
Paciente #4	75.00	68.54	84.44
Paciente #5	96.43	100.00	78.89
Paciente #6	78.57	92.31	72.22
Paciente #7	96.42	91.13	75.56
Paciente #8	93.47	86.39	77.67
Paciente #9	89.28	93.08	77.78
Paciente #10	98.25	71.54	82.14
Paciente promedio	94.73	75.57	93.68

En la tabla 4.18 se muestran las tasas de falsos positivos porcentual al ejecutar el ASN con los datos pertenecientes a los 11 pacientes simulados en los momentos en los que ocurre la falla del mediodía y la falla de las 4 de la tarde. También se observan los resultados obtenidos al ejecutar el ASN cuando no se presenta ninguna falla. La falla de medianoche no se toma en cuenta debido a que se está considerando un período de 24 horas para las simulaciones, el cual comienza justamente con esta falla.

Tabla 4. 18 Tasa de falsos positivos porcentuales aplicando ASN para cada paciente simulado (%).

Caso de estudio	Sin FEII	FEII de mediodía (12:00 p.m.)	FEII 4 de la tarde (4:00 p.m.)
Paciente #1	0.00	28.57	21.05
Paciente #2	7.14	7.14	5.26
Paciente #3	10.71	14.29	21.05
Paciente #4	25.00	0.00	15.79
Paciente #5	13.57	7.14	26.32
Paciente #6	21.43	14.29	15.79
Paciente #7	3.57	7.14	15.79
Paciente #8	0.00	28.57	14.89
Paciente #9	10.71	14.29	21.05
Paciente #10	0.00	7.14	5.26
Paciente promedio	5.26	7.14	26.32

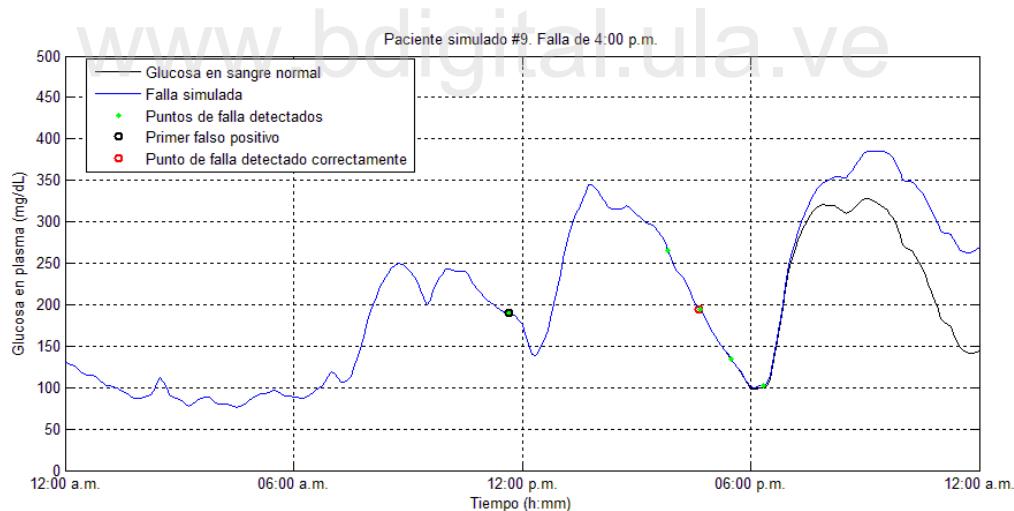


Figura 4. 7 Valores de glucosa en plasma paciente simulado #9 FEII 4:00 p.m. al utilizar ASN.

En la figura 4.7 se observan las curvas correspondientes a valores normales de glucosa y valores cuando se simula la falla de las 4:00 p.m. para el caso del paciente #9. La detección correcta del primer punto de falla se produce 40 minutos después de haberse

comenzado a simular la falla. Se puede observar que un período de 16 horas se detectan erróneamente 2 puntos como puntos de falla (falsos positivos).

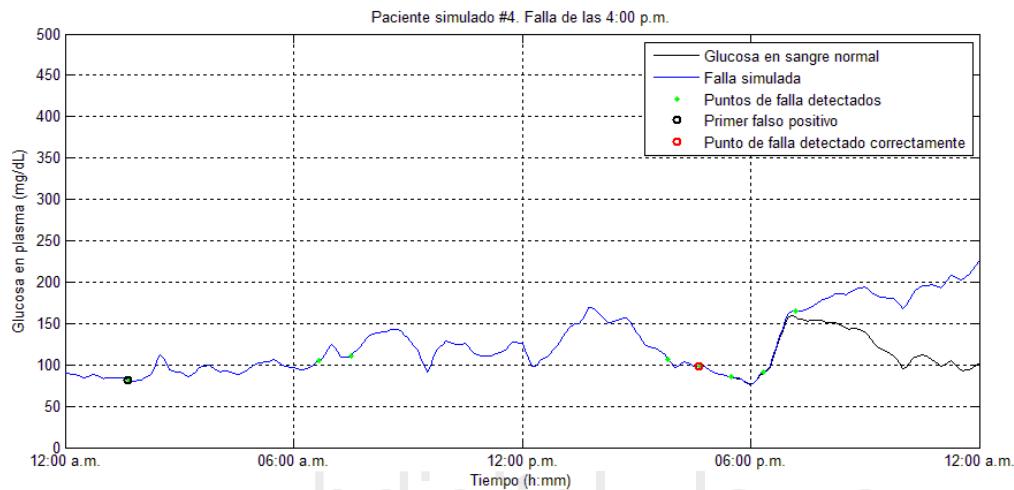


Figura 4.8 Valores de glucosa en plasma paciente simulado #4. Falla 4:00 p.m.

En la figura 4.8 se presentan las curvas correspondientes a los valores de glucosa en plasma del paciente simulado #4 en momentos en los que no se presenta FEII y cuando se presenta la falla de las 4:00 p.m. Como se observa, el primer punto de falla se detecta correctamente 30 minutos después de haberse comenzado a simular la falla. Se observa que se detectan erróneamente cuatro puntos de falla antes de que la misma empiece a simularse.

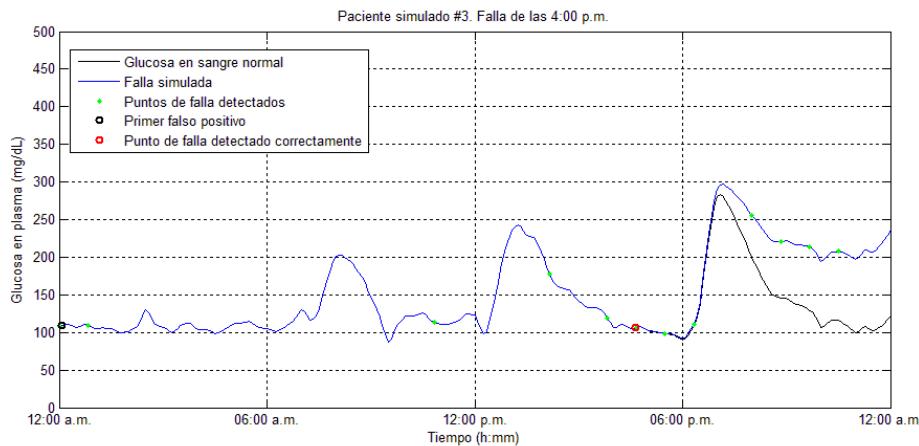


Figura 4. 9 Glucosa en plasma. Paciente simulado #3 Falla de las 4:00 p.m.

En la figura 4.9 se presentan las curvas correspondientes a valores de glucosa en plasma normales y valores cuando se presenta la FEII de las 4:00 p.m. en el paciente simulado #3. Como se observa ocurren un total de cinco detecciones anteriores a la falla, que corresponden a falsos positivos en un período de 16 horas de simulación. De igual manera, se aprecia que la detección del primer punto de falla correctamente, ocurre 30 minutos luego de haberse comenzado a simular la falla.

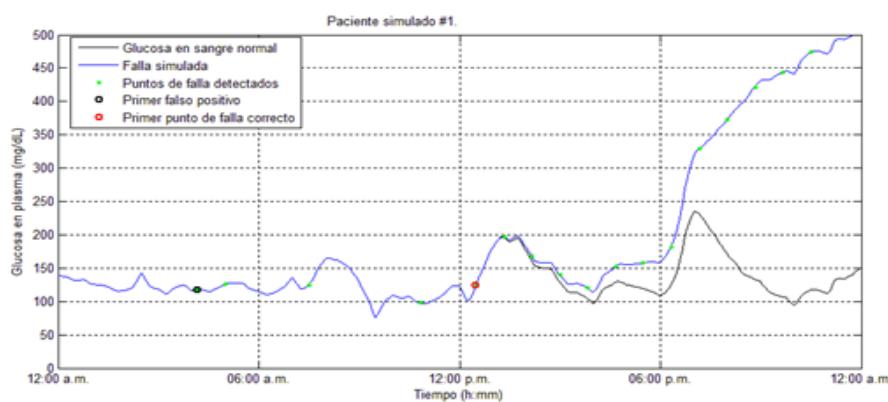


Figura 4. 10 Glucosa en plasma. Paciente simulado #1 Falla de las 12:00 p.m

En la figura 4.10 se presentan las curvas correspondientes a los valores normales de glucosa en plasma y los valores correspondientes a las concentraciones de glucosa cuando se simula la FEII de las 12:00 p.m. en el paciente #1. Se puede observar que el primer punto de falla se detecta 20 minutos después del comienzo de la simulación de la falla. De igual manera, se puede observar que ocurren cuatro falsos positivos en un período de 12 horas de simulación.

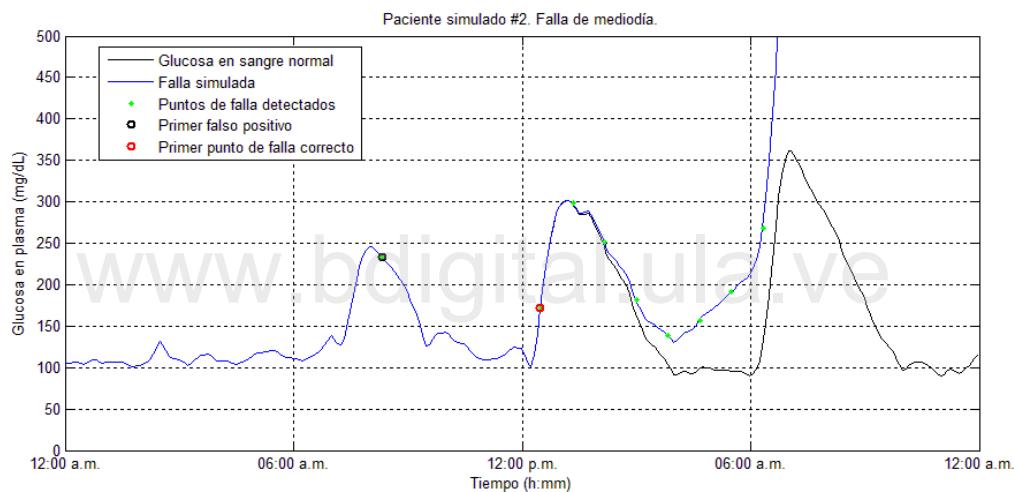


Figura 4. 11Glucosa en plasma. Paciente simulado #2 Falla de las 12:00 p.m

En la figura 4.11 se presentan las curvas correspondientes a los valores normales de glucosa en plasma y los valores correspondientes a las concentraciones de glucosa cuando se simula la FEII de las 12:00 p.m. en el paciente #2. Se puede observar que el primer punto de falla se detecta 30 minutos después del comienzo de la simulación de la falla. De igual manera, se puede observar que ocurre un falso positivo en un período de 12 horas de simulación

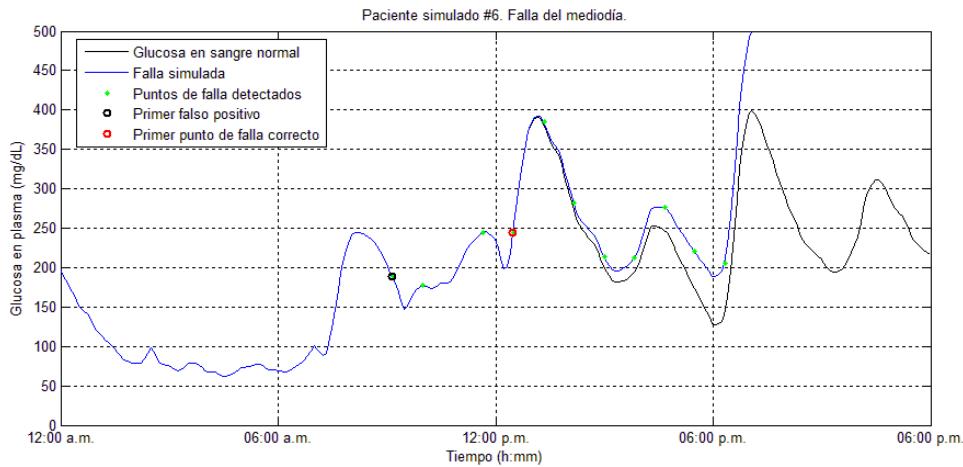


Figura 4. 12 Glucosa en plasma. Paciente simulado #6. Falla del mediodía.

En la figura 4.12 se presentan las curvas correspondientes a los valores normales de glucosa en plasma y los valores correspondientes a las concentraciones de glucosa cuando se simula la FEII de las 12:00 p.m. para el paciente #6. Se puede observar que el primer punto de falla se detecta 30 minutos después del comienzo de la simulación de la falla. De igual manera, se puede observar que ocurren tres falsos positivos en un período de 12 horas de simulación.

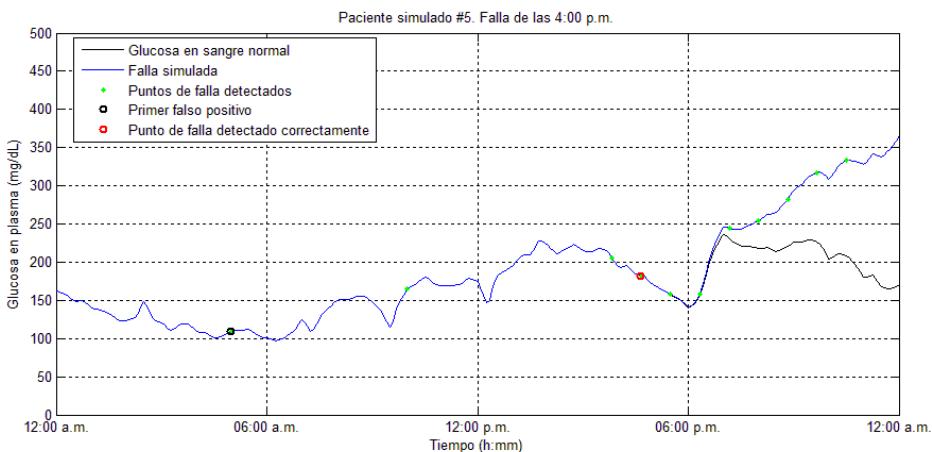


Figura 4. 13Glucosa en plasma paciente simulado #5. Falla de las 4:00 p.m.

En la figura 4.13 se presentan las curvas correspondientes a los valores normales de glucosa en plasma y los valores correspondientes a las concentraciones de glucosa cuando se simula la FEII de las 4:00 p.m. para el paciente #5. Se puede observar que el primer punto de falla se detecta 40 minutos después del comienzo de la simulación de la falla. También, se puede observar que ocurren tres falsos positivos en un período de 12 horas de simulación.

### 4.3 Comparación entre ASN y Enfoque Estadístico Multivariable

A continuación se presentan dos tablas resumen en las que se muestran los resultados obtenidos por Rojas et al. en la detección de FEII utilizando tres métodos estadísticos y los resultados obtenidos utilizando el ASN con detectores de radio variable. Para comparar los dos enfoques (estadístico e inmune) se toman en cuenta la tasa de falsos positivos por hora y los tiempos de detección.

Tabla 4. 19 Tasa de falsos positivos (# fp/hora, Media (Desviación Estándar)).

<i>Caso Experimental</i>	<i>Algoritmo ACP</i>	<i>Algoritmo combinado</i>	<i>Algoritmo CB</i>	<i>ASN (100% datos propios)</i>	<i>ASN (75% datos propios)</i>
Sin falla presente	0.04 (0.05)	0.08 (0.08)	0.21 (0.13)	0.00 (0.00)	0.23 (0.09)
Falla de mediodía	0.00 (0.00)	0.05 (0.08)	0.13 (0.06)	0.00 (0.00)	0.45 (0.24)
Falla de 4 de la tarde	0.02 (0.03)	0.07 (0.08)	0.02 (0.04)	0.00 (0.00)	0.33 (0.11)

En la tabla 4.19 se muestran las tasas de falsos positivos por hora obtenidas al ejecutar los algoritmos basados en un enfoque estadístico multivariable y el ASN de radio variable. Se

observa que al ejecutar el ASN con el conjunto total de datos esta tasa se reduce a cero. Pero, utilizando un porcentaje del conjunto de datos propios (75%), las tasas de falsos positivos son mayores a las reportadas por Rojas, et al. (2011).

Tabla 4. 20Tiempo de detección (HH:MM, Media [Desviación Estándar]).

<i>Caso Experimental</i>	<i>Algoritmo CB</i>	<i>Algoritmo combinado</i>	<i>Algoritmo PCA</i>	<i>ASN detectores de tamaño variable</i>
Falla de medianoche (12:00 a.m.)	7:26 (0:50)	6:04 (2:03)	6.04 [2:03]	1:21[0:25]
Falla de mediodía (4:00 p..m.)	1:04 (0:11)	2:05 (2:00)	3:04 [2:24]	0:30 [0:07]
Falla de 4 de la tarde (4:00 p.m.)	2:45 (0:51)	3:09 (1:35)	3:49 [0.48]	1:47 [0:35]

En la tabla 4.20 se presentan los resultados correspondientes a los tiempos de detección obtenidos al aplicar algoritmos basados en el enfoque estadístico multivariable y el algoritmo inmune implementado en nuestra investigación. Se puede observar que el ASN ofrece los menores tiempos de detección. En el caso de la falla simulada a medianoche la detección correcta del primer punto de falla ocurre, en promedio, una hora y 21 minutos después de haber comenzado la simulación de la FEII. En cuanto a la falla de mediodía, la detección ocurre, en promedio, 30 minutos después de haberse detectado la falla y en el caso de la falla de las 4:00 p.m. la media del tiempo de detección es de 1 hora 47 minutos. Esto representa una clara ventaja sobre los algoritmos basados en enfoques estadísticos, los cuales arrojan medias de tiempos de detección mayores para las tres fallas simuladas.

Es importante mencionar que el hecho de que las tasas de falsos positivos son mayores, cuando se ejecuta el ASN con 75% de los datos del conjunto propio (ver tabla 4.19), constituye una desventaja de este algoritmo con respecto a los algoritmos basados en enfoques estadísticos, ya que antes de producirse la primera detección correcta utilizando el ASN, se detectan equivocadamente datos normales como falla.

[www.bdigital.ula.ve](http://www.bdigital.ula.ve)

C.C. Reconocimiento

## Conclusiones

El propósito de esta investigación fue estudiar el desempeño del algoritmo de selección negativa de valores reales en la detección temprana de fallas de equipos de infusión de insulina que son utilizados por pacientes que padecen diabetes tipo I. La investigación se enfocó en realizar pruebas preliminares a dos versiones del ASN y evaluar su desempeño utilizando para ello un conjunto de datos apropiado. Se demostró experimentalmente, utilizando un conjunto de datos ampliamente utilizados para probar algoritmos de clasificación, que el ASN de valores reales con detectores de tamaño variable representa la implementación más eficiente. Al utilizar el ASN con detectores de radio fijo, se pudo observar que las tasas de falsos positivos son mayores, mientras que en el caso del ASN con detectores de radio variable es posible reducir esta tasa a cero. La cantidad de detectores generados en la versión de radio variable es mucho menor a la generada en la versión de radio fijo, lo cual aumenta la velocidad de ejecución del algoritmo. La implementación propuesta en nuestro trabajo fue comparada con una investigación publicada por Zhou, Ji et al. (2004), y se pudo determinar que es válida ya que cumple con el criterio de error absoluto porcentual entre las tasas de detección y de falsos positivos publicadas y obtenidas, inferior a 10%.

Luego de validar el ASN y determinar que la versión que ofrece mayores ventajas es la versión de radio variable, se procedió a utilizar el ASN con detectores de radio variable en la

detección de FEII en el grupo de pacientes *in silico*. Para este propósito se ajustaron valores de parámetros y se hizo un pre-procesamiento a datos obtenidos del simulador de pacientes para poder ser utilizados por el ASN. Finalmente, se compararon los resultados obtenidos con los resultados de otras investigaciones orientadas a la detección de FEII, pero que utilizan algoritmos basados en enfoques estadísticos. Se pudo demostrar que el ASN con detectores de radio variable ofrece una ventaja que tiene que ver con el tiempo de detección. Resultados publicados, en los que se utiliza el enfoque estadístico multivariable para la detección de FEII, muestran que la media de tiempos de detección es mayor, para todos los casos estudiados, que la media de tiempos de detección obtenida al ejecutar el ASN. Recordemos que, en este caso, es de gran importancia que la detección de la falla ocurra en un tiempo adecuado para que se tomen acciones correctivas antes de que las concentraciones de glucosa en sangre alcancen valores que pongan en riesgo al paciente que utiliza este tipo de sistemas de infusión continua de insulina. La tasa de falsos positivos por hora y el tiempo de detección son los dos puntos clave tomados en cuenta para realizar la comparación entre los algoritmos presentados. En tal sentido, se determinó que los algoritmos inmunes ofrecen ventajas en cuanto al tiempo de detección, pero en cuanto a las tasas de falsos positivos por hora, los algoritmos basados en enfoques estadísticos, ofrecen tasas menores. El propósito de la implementación de estos algoritmos, es obviamente la detección de las fallas, es decir, que sean eficientes al momento de clasificar correctamente los datos como normales o anormales, pero si esta tarea no es realizada en el tiempo requerido, resulta poco efectiva. Es por eso que se consideran los algoritmos inmunes como una herramienta con gran potencial en el campo de investigación relacionado con la detección de fallas en equipos de infusión de insulina.

Este estudio puede ser considerado como un aporte a investigaciones orientadas a la detección de este tipo de fallas, ya que no existe evidencia publicada de que el ASN haya sido utilizado para este propósito. El área de investigación relacionada con los sistemas inmunes artificiales y con la detección de FEII sigue siendo objeto de estudio y en futuras investigaciones podrían emplearse el ASN con reglas de coincidencia distintas a la distancia Euclidiana para realizar la clasificación de los datos. De igual manera, es conveniente realizar pruebas con los otros grupos de pacientes incluidos en el simulador UVA/PADOVA. Otro aspecto que podría incluirse en futuros trabajos es utilizar distintas técnicas para el pre-procesamiento de los datos que van a ser utilizados y finalmente, se recomienda hacer pruebas con datos obtenidos de pacientes reales.

[www.bdigital.ula.ve](http://www.bdigital.ula.ve)

C.C. Reconocimiento

## Referencias Bibliográficas

- [1] Araujo, M., Aguilar, J., & Aponte, H. (2003, July). Fault detection system in gas lift well based on Artificial Immune System. In *Proceedings of the International Joint Conference on Neural Networks, 2003*. (Vol. 3, pp. 1673-1677). IEEE.
- [2] Baysal, N., Cameron, F., Buckingham, B. A., Wilson, D. M., & Bequette, B. W. (2013, June). Detecting sensor and insulin infusion set anomalies in an artificial pancreas. In *2013 American Control Conference* (pp. 2929-2933). IEEE.
- [3] Bequette, B. W. (2005). A critical assessment of algorithms and challenges in the development of a closed-loop artificial pancreas. *Diabetes technology & therapeutics*, 7(1), 28-47.
- [4] Blake, E. K., Merz, C., "UCI Repository of Machine Learning Databases", Irvine, CA. University of California, Department of Information and Computer Science, 1998. (Last accessed Jun 7, 2009).
- [5] Bradley, D. W., & Tyrrell, A. M. (2000, April). Immunotronics: Hardware fault tolerance inspired by the immune system. In *International Conference on Evolvable Systems* (pp. 11-20). Springer, Berlin, Heidelberg.
- [6] Cescon, M., DeSalvo, D. J., Ly, T. T., Maahs, D. M., Messer, L. H., Buckingham, B. A & Dassau, E. (2016). Early detection of infusion set failure during insulin pump therapy in type 1 diabetes. *Journal of diabetes science and technology*, 10(6), 1268-1276.
- [7] Dasgupta, D., & Forrest, S. (1996, June). Novelty detection in time series data using ideas from immunology. In *Proceedings of the international conference on intelligent systems* (pp. 82-87).

- [8] Dasgupta, D., & Forrest, S. (1999, July). Artificial immune systems in industrial applications. In *Intelligent Processing and Manufacturing of Materials, 1999. IPMM'99. Proceedings of the Second International Conference on* (Vol. 1, pp. 257-267). IEEE.
- [9] De Castro, L. N., & Timmis, J. I.(2002). Artificial immune systems: a novel paradigm to pattern recognition, *Artificial Neural Networks in Pattern Recognition*, JM Corchado, L. Alonso, C. Fyfe (University of Paisley, UK, 2002), 67-84.
- [10] Diabetes Tipo 1. © 1998-2020 Mayo Foundation for Medical Education and Research. <https://www.mayoclinic.org/es-es/diseases-conditions/type-1-diabetes/symptoms-causes/syc-20353011>
- [11] García A., Diego (2009). Fault detection using principal component analysis (PCA) in a wastewater treatment plant (WWTP).
- [12] González, F. A., & Dasgupta, D. (2003). Anomaly detection using real-valued negative selection. *Genetic Programming and Evolvable Machines*, 4(4), 383-403.
- [13] Guilhem, I., Leguerrier, A. M., Lecordier, F., Poirier, J. Y., & Maugendre, D. (2006). Technical risks with subcutaneous insulin infusion. *Diabetes & metabolism*, 32(3), 279-284.
- [14] Herrero, P., Calm, R., Vehí, J., Armengol, J., Georgiou, P., Oliver, N., & Tomazou, C. (2012). Robust fault detection system for insulin pump therapy using continuous glucose monitoring. *Journal of diabetes science and technology*, 6(5), 1131-1141.
- [15] Ji, Z., & Dasgupta, D. (2004, June). Real-valued negative selection algorithm with variable-sized detectors. In *Genetic and Evolutionary Computation Conference* (pp. 287-298). Springer, Berlin, Heidelberg.
- [16] Ji, Z. (2006). *Negative Selection Algorithms: from the Thymus to V-detector*. The University of Memphis.

- [17] Kölle, K., Fougner, A. L., Unstad, K. A. F., & Stavdahl, Ø. (2018). Fault detection in glucose control: Is it time to move beyond CGM data?. *IFAC-PapersOnLine*, 51(27), 180-185.
- [18] Kovatchev BP, Breton M, Dalla Man C, Cobelli C. In silico preclinical trials: a proof of concept in closed-loop control of type 1 diabetes. *J Diabetes Science Technol.* 2009; 3(1):44-55.
- [19] Li T. Computer immunology. Beijing Publishing House of Electronics Industry; 2004.
- [20] Mohammadi, M., Akbari, A., Raahemi, B., Nassersharif, B., & Asgharian, H. (2014). A fast anomaly detection system using probabilistic artificial immune algorithm capable of learning new attacks. *Evolutionary Intelligence*, 6(3), 135-156.
- [21] Nino, F., & Dasgupta, D. (2008). *Immunological computation: theory and applications*. Auerbach Publications.
- [22] Rojas, R., Garcia-Gabin, W., & Bequette, B. W. (2011, January). Mean glucose slope principal component analysis classification to detect insulin infusion set failure. In *Proceedings of the 18th IFAC World Congress, Milan, Italy* (Vol. 29).
- [23] Rojas, R., Garcia-Gabin, W., & Bequette, B. W. (2011, June). Multivariate statistical analysis to detect insulin infusion set failure. In *American Control Conference (ACC), 2011* (pp. 1952-1957). IEEE.
- [24] “Statlib datasets archive”, World Wide Web, <http://lib.stat.cmu/datasets>. (Last accessed May 12, 2009).

- [25] Visentin, R., Campos-Náñez, E., Schiavon, M., Lv, D., Vettoretti, M., Breton, M., & Cobelli, C. (2018). The UVA/Padova type 1 diabetes simulator goes from single meal to single day. *Journal of diabetes science and technology*, 12(2), 273-281.

[www.bdigital.ula.ve](http://www.bdigital.ula.ve)

C.C. Reconocimiento