



**INAOE**

# **Cuantificación de hemoglobina glicosilada y glucosa *in vivo* usando espectroscopia Raman y redes neuronales artificiales**

Por:

**M. I. Naara González Viveros**

Tesis sometida como requisito parcial  
para obtener el grado de:

**DOCTORA EN CIENCIAS EN LA ESPECIALIDAD DE  
ÓPTICA**

en el

**Instituto Nacional de Astrofísica, Óptica y Electrónica**

Junio, 2021

Tonantzintla, Puebla, México

Dirigida por:

**Dr. Jorge Castro Ramos, INAOE**

**Dra. María del Pilar Gómez Gil, INAOE**

**Dr. Hector Hugo Cerecedo Núñez, UV**

©INAOE 2021

Derechos reservados

El autor otorga al INAOE el permiso de  
reproducir y distribuir copias de esta tesis en su  
totalidad o en partes mencionando la fuente.





---

## DEDICATORIA

---

*A mi hermano **Joel González** que descanse en paz  
(1982-2018).*

*A mis amados padres **Joel González y Eva Viveros** por  
su ejemplo, dedicación y por siempre creer en mí.*

*A mi hermano **Uriel González**, por ser mi más grande  
motivación para continuar con esta investigación.*

*A mis sobrinos: **Donaji, Joel, Nikte y Jordan.***

*A mi amado **Alex Zenteno** por todo su apoyo, amor y  
comprensión.*



---

## AGRADECIMIENTOS

---

**Al Consejo Nacional de Ciencia y Tecnología (CONACyT)**

Por el apoyo económico otorgado a través de la beca para estudios de doctorado.

**A mis asesores:**

**Dr. Jorge Castro Ramos**

**Dra. María del Pilar Gómez Gil**

**Dr. Héctor Hugo Cerecedo Núñez.**

Por su paciencia y dedicación en la asesoría de esta tesis, por su confianza y conocimientos, gracias infinitas.

**Al laboratorio Diagnoscentro, al Centro de Estudios y Prevención del Cáncer (CEPREC) y a la Clínica 2 IMSS- Puebla**

por proporcionar las muestras necesarias para el desarrollo de este trabajo.

**Al Grupo de Instrumentación y Óptica Biomédica (GIOB)**

Por el apoyo, comentarios y la convivencia durante estos años.

**A mis sinodales:**

**Dr. Francisco Renero, Dr. Alberto Jaramillo, Dr. Juan Manuel Ramírez, Dr. Vicente Alarcón, Dr. Francisco Gutiérrez.**

Por sus comentarios y tomarse el tiempo para leer este trabajo.

Así mismo al **Msc José Luis Flores Guerrero, Dr. Jonás Grande Barreto, Dr. Saúl Solorio Fernandez**, gracias por compartir sus conocimientos y consejos para la realización de este trabajo.

Agradezco a **Claudia Cruz, Ricardo Jiménez, Andrea Burgos, Jesús García, Eréndira Vázquez, Daniela Reynoso**, entre muchas otras personas con las cuales compartí momentos que siempre recordaré. También gracias a mis amigos **Francisco Molina y Rosa M. Rodríguez**, porque a pesar de los kilómetros que nos separan siempre me apoyaron.



---

## ABSTRACT

---

This research introduces a novel model for the quantification of glycated hemoglobin (HbA<sub>1c</sub>) and glucose in diabetic and non-diabetic patients, which is based on Raman spectroscopy as well as in classification and regression techniques using artificial neural networks (ANN). Our proposal was tested in 42 diabetics and 4 non-diabetic volunteers using Raman spectra signals obtained from their forearms, wrists, index fingers, and blood serum. Base values of HbA<sub>1c</sub> and glucose were obtained from standard laboratory tests; the analyzed HbA<sub>1c</sub> values ranged from 5.2 % to 14 % and glucose values ranged from 56 to 400 mg/dL.

Different methods for feature selection as well as feature extraction were analyzed, including RReliefF (Regression ReliefF), CSF (Correlation Selection Feature), Wrapper, and PCA (Principal Component Analysis), in order to find a combination producing the lowest Root Mean Square Error in Cross-Validation (RMSE-CV). Indeed, we propose a feature enhancement method based on Kohonen's self-organizing maps (SOM) applied to Raman spectra, showing an improvement in calculated glucose and HbA<sub>1c</sub> values.

In addition, spectroscopic characterization of pure HbA<sub>1c</sub> and glucose was carried out, locating representative Raman peaks; the molecular assignment was made according to the reported in the literature. These peaks were consistent with the ones found in the *in vivo* and *in vitro* spectra obtained from volunteers.

As overall, we obtained an RMSE-CV of **11.86 ± 5.45 mg/dL** using a model based on RReliefF and SOM network when quantifying glucose in serum, which is better compared to works reported in the state of the art, when methods of feature selection were not used. Regarding the *in vivo* measurements, an error of **0.69 % ± 0.07** was obtained to determine the HbA<sub>1c</sub> concentrations in volunteers. It should be noted that no work has been reported nowadays in which the percentage of HbA<sub>1c</sub> in humans has been directly quantified;

therefore, this is the first one. With respect to the measurements of glucose *in vivo*, a RMSE-CV of  $30.12 \pm 0.53$  mg/dL was obtained in measurements of the forearm, showing a percentage of success in the Clarke error grid of **82.61** % in zone A, **15.22** % in zone B, **2.17** % in zone D and **0** % for zone C and E. According to these results, we conclude that by using Raman spectroscopy, feature selection methods, and artificial neural network is possible to quantify HbA<sub>1c</sub> and glucose *in vivo* with an error comparable and smaller in some experiments than the error reported in similar works in state of the art nowadays.



---

## RESUMEN

---

Esta investigación presenta un modelo novedoso para la cuantificación de hemoglobina glicosilada (HbA<sub>1c</sub>) y glucosa en pacientes diabéticos y no diabéticos, el cual se basa en espectroscopia Raman, así como en técnicas de clasificación y regresión mediante redes neuronales artificiales (ANN). Esta propuesta se probó en 42 voluntarios diabéticos y 4 no diabéticos utilizando las señales de espectros Raman obtenidas de sus antebrazos, muñecas, dedos índices y suero sanguíneo. Los valores base de HbA<sub>1c</sub> y glucosa se obtuvieron de pruebas estándar de laboratorio; los valores de HbA<sub>1c</sub> analizados oscilaron entre 5,2 y 14 % y los valores de glucosa oscilaron entre 56 y 400 mg/dL.

Se analizaron diferentes métodos para la selección de características, así como para la extracción de características, incluidos RReliefF (Regresión ReliefF), CSF (selección de característica por correlación), Wrapper (método de envoltura) y PCA (Análisis de Componentes Principales), con el fin de encontrar la combinación que produjera el menor error dado por la raíz del error cuadrático medio en validación cruzada (RMSE-CV, por sus siglas en inglés). También, se propone un método de mejoramiento de características basado en los mapas auto-organizados de Kohonen (SOM, por sus siglas en inglés) aplicados a los espectros Raman, que muestra una mejora en la predicción de valores de glucosa y HbA<sub>1c</sub>. Además, se realizó una caracterización espectroscópica de HbA<sub>1c</sub> pura y glucosa, en donde se localizaron los picos Raman representativos, así mismo se realizó la asignación molecular de acuerdo con los valores reportados en la literatura. Estos picos fueron coherentes con los encontrados en los espectros *in vivo* e *in vitro* obtenidos de voluntarios.

En general, se obtuvo un RMSE-CV de **11.86 ± 5.45 mg/dL** utilizando un modelo basado en RReliefF y red SOM al cuantificar la glucosa en suero, lo cual es mejor comparado con trabajos reportados en el estado del arte, en los cuales no utilizan métodos de selección de características. Con respecto a las

mediciones *in vivo* se obtuvo un error de **0.69 %  $\pm$  0.07** para determinar las concentraciones de HbA<sub>1c</sub> en voluntarios en las mediciones de la muñeca, cabe resaltar que hasta la fecha no se ha reportado algún trabajo en el estado del arte en el cual se cuantifique directamente el porcentaje de HbA<sub>1c</sub> en humanos, siendo este el primero.

Con respecto a las mediciones de glucosa *in vivo*, se obtuvo un RMSE-CV de **30.12  $\pm$  0.53 mg/dL** en las mediciones del antebrazo, mostrando un porcentaje de éxito en la rejilla de error de Clarke de **82.61 %** en la zona A, **15.22 %** en la zona B, **2.17 %** en la zona D y **0 %** para las zonas C y E.

De acuerdo con estos resultados, se concluye que mediante el uso de espectroscopia Raman, métodos de selección de características y redes neuronal artificial es posible cuantificar HbA<sub>1c</sub> y glucosa *in vivo* con un error comparable y en ocasiones menor que el error reportado en trabajos similares reportados en el estado del arte en la actualidad.

---

# TABLA DE CONTENIDOS

---

|  |            |
|--|------------|
| <b>Abstract</b>  | <b>VII</b> |
| <b>Resumen</b>   | <b>IX</b>  |
| <b>1. Introducción</b>   | <b>1</b>   |
| 1.1. Estado del Arte . . . . .   | 3          |
| 1.2. Hipótesis . . . . .   | 7          |
| 1.3. Objetivos . . . . .   | 7          |
| 1.3.1. Objetivo general . . . . .  | 7          |
| 1.3.2. Objetivos particulares . . . . .                                  | 7          |
| 1.4. Resumen del documento . . . . .                                     | 8          |
| <b>2. Marco Teórico</b>  | <b>9</b>   |
| 2.1. Espectroscopia Raman . . . . .                                      | 9          |
| 2.1.1. Efecto Raman . . . . .  | 9          |
| 2.1.2. Tipos de ruido en espectroscopia Raman . . . . .                  | 14         |
| 2.1.3. Eliminación de fluorescencia y suavizado . . . . .                | 16         |
| 2.2. Aprendizaje de máquina . . . . .                                    | 17         |
| 2.2.1. Redes Neuronales Artificiales . . . . .                           | 19         |
| 2.2.2. Máquina de Soporte Vectorial . . . . .                            | 23         |
| 2.2.3. Regresión lineal . . . . .  | 24         |
| 2.2.4. Regresión de mínimos cuadrados parciales por intervalos . . . . . | 25         |
| 2.3. Procesamiento de datos . . . . .                                    | 25         |
| 2.3.1. Mapas Auto-organizados . . . . .                                  | 26         |
| 2.3.2. Extracción de características . . . . .                           | 27         |
| 2.3.3. Selección de características . . . . .                            | 29         |
| 2.4. Resumen del capítulo . . . . .                                      | 34         |

---

|   |           |
|---|-----------|
| <b>3. Desarrollo Experimental</b>                                   | <b>35</b> |
| 3.1. Mediciones de Glucosa . . . . .                                | 35        |
| 3.2. Mediciones de HbA <sub>1c</sub> . . . . .                      | 36        |
| 3.3. Mediciones de suero humano . . . . .                           | 37        |
| 3.4. Mediciones <i>in vivo</i> . . . . .                            | 39        |
| 3.4.1. Selección de parámetros . . . . .                            | 39        |
| 3.4.2. Selección de regiones corporales . . . . .                   | 42        |
| 3.4.3. Adquisición de espectros Raman <i>in vivo</i> . . . . .      | 44        |
| 3.5. Resumen del capítulo . . . . .                                 | 46        |
| <br>  |           |
| <b>4. Procesos de Medición</b>                                      | <b>47</b> |
| 4.1. Regresión . . . . .  | 48        |
| 4.1.1. Mapas auto-organizados (SOM) . . . . .                       | 50        |
| 4.1.2. Análisis de Componentes Principales . . . . .                | 51        |
| 4.1.3. Selección de características basada en correlación . . . . . | 52        |
| 4.1.4. Selección de características usando RReliefF . . . . .       | 53        |
| 4.1.5. Selección de características usando Wrapper . . . . .        | 54        |
| 4.2. Otros métodos utilizados . . . . .                             | 56        |
| 4.3. Clasificación . . . . .  | 56        |
| 4.4. Análisis espectral . . . . .                                   | 57        |
| 4.5. Resumen del capítulo . . . . .                                 | 58        |
| <br>  |           |
| <b>5. Resultados y discusión</b>                                    | <b>59</b> |
| 5.1. Base de datos mediciones <i>in vivo</i> . . . . .              | 59        |
| 5.2. Regresión . . . . .  | 62        |
| 5.2.1. Mapas auto-organizados . . . . .                             | 62        |
| 5.2.2. Análisis de Componentes Principales . . . . .                | 65        |
| 5.2.3. Selección de características basada en correlación . . . . . | 68        |
| 5.2.4. RReliefF . . . . .   | 69        |
| 5.2.5. Wrapper . . . . .  | 72        |
| 5.3. Otros métodos utilizados . . . . .                             | 74        |
| 5.4. Clasificación . . . . .  | 79        |
| 5.5. Análisis espectral . . . . .                                   | 82        |
| 5.6. Resumen de capítulo . . . . .                                  | 86        |

|  |           |
|--|-----------|
| <b>6. Conclusiones y trabajo futuro</b>  | <b>87</b> |
| 6.1. Contribuciones . . . . .  | 89        |
| 6.2. Trabajo futuro . . . . .  | 90        |
| 6.3. Trabajos y participaciones derivadas de esta tesis . . . . .                                  | 91        |
| <b>References</b>  | <b>93</b> |
| <b>A. Figuras y tablas complementarias</b>   | <b>10</b> |
| A.1. Espectros de mediciones <i>in vivo</i> . . . . .  | 10        |
| A.2. Análisis de k-vecinos y características por región . . . . .                                  | 11        |
| A.3. Rejilla de error de Clarke por región . . . . .   | 12        |
| A.4. Análisis espectral mediciones <i>in vivo</i> . . . . .  | 13        |
| <b>B. Programas</b>  | <b>22</b> |
| B.1. Rejilla de error de Clarke . . . . .  | 22        |
| B.2. RReliefF-SOM . . . . .  | 25        |
| B.3. CFS, Wrapper-SOM . . . . .  | 30        |
| B.4. PCA-SOM . . . . .   | 35        |
| B.5. SVM-LR . . . . .  | 40        |
| B.6. Clasificación . . . . .   | 42        |
| <b>C. Consentimiento informado</b>   | <b>44</b> |
| <b>D. Algunos métodos médicos</b>  | <b>45</b> |
| D.1. Rejilla de Error de Clarke . . . . .  | 45        |
| D.2. Método de glucosa oxidasa . . . . .   | 46        |
| D.3. Método de afinidad por boronato . . . . .   | 47        |
| <b>E. Hiper-parámetros de los algoritmos de entrenamiento de las redes neuronales artificiales</b> | <b>48</b> |
| E.1. Levenberg-Marquardt . . . . .   | 48        |
| E.2. Scaled Conjugate Gradient . . . . .   | 48        |



---

## LISTA DE FIGURAS

---

|       |  |    |
|-------|--|----|
| 1.1.  | <i>Fórmula química de la HbA<sub>1c</sub> [1, 2]</i>   | 2  |
| 2.1.  | <i>Experimento de Chandrasekhara Venkata Raman [3].</i>  | 10 |
| 2.2.  | <i>Molécula diatómica con una masa en un resorte [4].</i>  | 11 |
| 2.3.  | <i>Diagrama de Jablonski que representa transiciones de energía cuántica para el esparcimiento de Rayleigh y Raman [5, 6].</i>   | 13 |
| 2.4.  | <i>Ejemplo de ruidos en el espectros Raman: a) Ruido de disparo, b) Bandas Raman, c) Ruido generado por la muestra (fluorescencia), d) Ruido generado por fuentes externas (rayos cósmicos).</i> | 16 |
| 2.5.  | <i>Representación gráfica de una validación cruzada de 5 particiones.</i>  | 18 |
| 2.6.  | <i>Gráfico arquitectónico de un perceptrón multicapa o FFNN con dos capas ocultas [7].</i>   | 20 |
| 2.7.  | <i>Estructura básica de la unidad fundamental de una red neuronal [7].</i>   | 20 |
| 2.8.  | <i>Funciones de activación a) Lineal, b) Sigmoidea tangente hiperbólica [8, 9].</i>  | 21 |
| 2.9.  | <i>Matriz de confusión 3 clases [10].</i>  | 22 |
| 2.10. | <i>Modelo de mapa topológico de Kohonen con la unidad de mejor coincidencia en verde, las capas dentro del radio de vecindad en morado y los nodos en el exterior en azul [7, 11].</i>           | 26 |
| 2.11. | <i>Enfoque de envoltura para la selección de subconjuntos de características. El algoritmo de inducción se utiliza como una “caja negra” por el algoritmo de selección de subconjuntos [12].</i> | 33 |
| 3.1.  | <i>Configuración experimental y análisis de datos.</i>   | 36 |
| 3.2.  | <i>Configuración experimental: a) Espectrómetro Raman QE65000, b) Fuente láser de 785 nm, c) Computadora, d) Punta de prueba Raman, e) Recipiente de muestra.</i>                                | 38 |

---

|      |   |    |
|------|---|----|
| 3.3. | <i>Prototipo para mediciones in vivo.</i>   | 40 |
| 3.4. | <i>Proceso de mediciones in vivo mediante espectroscopia Raman.</i>   | 45 |
| 4.1. | <i>Diseño de sistemas de regresión y clasificación para la determinación de HbA<sub>1c</sub> y glucosa.</i> | 48 |
| 4.2. | <i>Sistema de medición utilizando mapas auto-organizados SOM.</i>   | 51 |
| 4.3. | <i>Sistema de medición utilizando Análisis de Componentes Principales y SOM.</i>                            | 52 |
| 4.4. | <i>Sistema de medición utilizando selección de características basada en correlación (CFS).</i>             | 53 |
| 4.5. | <i>Sistema de medición utilizando RReliefF.</i>   | 54 |
| 4.6. | <i>Sistema de medición utilizando Wrapper.</i>  | 55 |
| 4.7. | <i>Sistema de medición utilizando un clasificador.</i>  | 57 |
| 5.1. | <i>Espectros Raman promedio por voluntario en la región de antebrazo.</i>                                   | 60 |
| 5.2. | <i>Espectro Raman de un voluntario con 5.2 % de HbA<sub>1c</sub>.</i>                                       | 63 |
| 5.3. | <i>Prototipo Red SOM del espectro Raman del voluntario con 5.2 % de HbA<sub>1c</sub>.</i>                   | 63 |
| 5.4. | <i>Ejemplo variación de K-Vecinos por número de características en la región del antebrazo.</i>             | 70 |
| 5.5. | <i>RMSE-CV por parte del cuerpo y número de características usando RReliefF.</i>                            | 71 |
| 5.6. | <i>Rejilla de error de Clarke para las predicción de glucosa en suero.</i>                                  | 78 |
| 5.7. | <i>Picos Raman de HbA<sub>1c</sub> pura comparados con 14 % y 5.2 % in vivo.</i>                            | 84 |
| 5.8. | <i>Picos Raman de glucosa pura comparados con 56 mg/dL y 400 mg/dL in vivo.</i>                             | 85 |
| A.1. | <i>Espectros Raman promedio por voluntario en la región de muñeca.</i>                                      | 10 |
| A.2. | <i>Espectros Raman promedio por voluntario en la región de dedo.</i>  | 10 |
| A.3. | <i>Espectros Raman promedio por voluntario en suero.</i>  | 11 |
| A.4. | <i>Variación de K-Vecinos por número de características en región de la muñeca.</i>                         | 11 |
| A.5. | <i>Variación de K-Vecinos por número de características en región del dedo.</i>                             | 12 |
| A.6. | <i>RMSE-CV predicción de glucosa por parte del cuerpo y número de características usando RReliefF.</i>      | 12 |

---



---

|  |    |
|--|----|
| A.7. <i>Rejilla de error de Clarke para las predicción de glucosa en antebrazo.</i> . .  | 13 |
| A.8. <i>Rejilla de error de Clarke para las predicción de glucosa en muñeca.</i> . .   | 13 |
| A.9. <i>Rejilla de error de Clarke para las predicción de glucosa en dedo.</i> . . . .   | 14 |
| C.1. <i>Consentimiento informado para mediciones no invasivas.</i> . . . . .   | 44 |
| D.1. <i>Rejilla de análisis de error para la evaluación de las implicaciones<br/>clínicas de los valores de glucosa[13].</i> . . . . . | 46 |



---

## LISTA DE TABLAS

---

|   |    |
|---|----|
| 2.1. Vibraciones típicas de algunas moléculas y enlaces. . . . .                                    | 15 |
| 3.1. Resultado del cálculo de máxima exposición permitida (MPE)<br>para diferentes tiempos. . . . . | 41 |
| 3.2. Relación señal a ruido (SNR) de los 21 voluntarios. . . . .                                    | 42 |
| 3.3. Análisis de diferencia de picos representativos de glucosa. . . . .                            | 43 |
| 3.4. Análisis de diferencia de picos representativos de HbA1c. . . . .                              | 43 |
| 3.5. Análisis de diferencia de picos representativos de glucosa. . . . .                            | 44 |
| 4.1. Representación de datos de entrada para WEKA. . . . .  | 52 |
| 5.1. Datos de voluntarios . . . . .   | 61 |
| 5.2. Resultado de la ANN con datos sin procesar. . . . .  | 62 |
| 5.3. Resultados de RMSE-CV para concentraciones de HbA1c usando<br>Red SOM. . . . .                 | 64 |
| 5.4. Resultados de RMSE-CV para glucosa usando Red SOM. . . . .                                     | 65 |
| 5.5. Resultados de RMSE-CV para HbA1c usando PCA . . . . .  | 66 |
| 5.6. Resultados de RMSE-CV para glucosa usando PCA. . . . .   | 66 |
| 5.7. Resultados de RMSE-CV para HbA1c usando PCA-SOM. . . . .                                       | 67 |
| 5.8. Resultados de RMSE-CV para glucosa usando PCA-SOM. . . . .                                     | 67 |
| 5.9. Resultados de RMSE-CV para HbA1c usando CFS-SOM. . . . .                                       | 69 |
| 5.10. Resultados de RMSE-CV para glucosa usando CFS-SOM. . . . .                                    | 69 |
| 5.11. Resultados de RMSE-CV para HbA1c usando RReliefF-SOM. . . . .                                 | 71 |
| 5.12. Resultados de RMSE-CV para glucosa usando RReliefF-SOM. . . . .                               | 72 |
| 5.13. Resultados de RMSE-CV para HbA1c usando Wrapper-SOM. . . . .                                  | 73 |
| 5.14. Resultados de RMSE-CV para glucosa usando Wrapper-SOM. . . . .                                | 73 |
| 5.15. Complejidad computación, basada en el tiempo de cómputo. . . . .                              | 75 |

|  |    |
|--|----|
| 5.16. Resultados de RMSE-CV para HbA1c comparación con SVM y LR. . . . .                                       | 76 |
| 5.17. Resultados de RMSE-CV para glucosa comparación con SVM y LR. . . . .                                     | 76 |
| 5.18. Resultados de RMSE-CV para HbA1c comparación iPLS y ANN. . . . .   | 76 |
| 5.19. Resultados de RMSE-CV para glucosa comparación iPLS y ANN. . . . .                                       | 77 |
| 5.20. Porcentaje de aciertos por zona en la rejilla de error de Clarke. . . . .                                | 77 |
| 5.21. Selección del porcentaje de datos para cada conjunto. . . . .  | 79 |
| 5.22. Resultados en porcentaje (%) de precisión para HbA1c en clasificación. . . . .                           | 80 |
| 5.23. Resultados en porcentaje (%) de precisión para HbA1c en clasificación usando red SOM. . . . .            | 80 |
| 5.24. Especificidad y sensibilidad mejor de los casos. . . . .   | 81 |
| 5.25. Comparación de métodos usados para mediciones <i>in vitro</i> de HbA1c. . . . .                          | 82 |
| 5.26. Comparación de métodos usados para mediciones <i>in vivo</i> de glucosa. . . . .                         | 82 |
| 5.27. Comparación de métodos usados para mediciones <i>in vitro</i> de glucosa. . . . .                        | 82 |
| 5.28. Asignación molecular de picos de HbA1c en polvo, basada en otros trabajos. . . . .                       | 83 |
| 5.29. Asignación molecular de picos de glucosa en polvo basada en otros trabajos. . . . .                      | 84 |
| 5.30. Relación señal a ruido (SNR) promedio de los 46 voluntarios. . . . .                                     | 86 |
| A.1. Picos localizados de HbA1c en muñeca. . . . .   | 15 |
| A.2. Picos localizados de HbA1c en antebrazo. . . . .  | 16 |
| A.3. Picos localizados de HbA1c en dedo. . . . .   | 17 |
| A.4. Picos localizados de glucosa en antebrazo. . . . .  | 18 |
| A.5. Picos localizados de glucosa en muñeca. . . . .   | 19 |
| A.6. Picos localizados de glucosa en dedo. . . . .   | 20 |
| A.7. Picos localizados de glucosa en suero. . . . .  | 21 |
| E.1. Parámetros algoritmo de entrenamiento Levenberg-Marquardt usado para los experimentos reportados. . . . . | 49 |

---

|   |    |
|---|----|
| E.2. Hiper-parámetros de entrenamiento usados en regresión . . . . .  | 49 |
| E.3. Parámetros algoritmo de entrenamiento gradiente conjugado<br>escalado usados para los experimentos reportados. . . . . | 49 |
| E.4. Hiper-parámetros de entrenamiento usados en clasificación. . . .   | 50 |



---

## GLOSARIO DE TÉRMINOS

---

HbA<sub>1c</sub>: Hemoglobina glicosilada, examen sanguíneo que mide el nivel promedio de glucosa o azúcar en la sangre durante los últimos tres meses.

Hb: Hemoglobina, proteína que se encuentra en los glóbulos rojo.

Mediciones *in vivo*: referente a las mediciones realizadas a un ser vivo.

Mediciones *in vitro*: referente a las mediciones que se realizan a nivel laboratorio, fuera de un organismo vivo.

Método de glucosa oxidasa: método para cuantificar niveles de glucosa en suero u orina.

Método de afinidad de boronato: método para cuantificar niveles de hemoglobina glicosilada en sangre.

Sangre sérica: fracción de la sangre que contiene el suero o plasma sanguíneo.

mg/dL: Unidad de medición de la glucosa (miligramos sobre decilitro)





---

## INTRODUCCIÓN

---

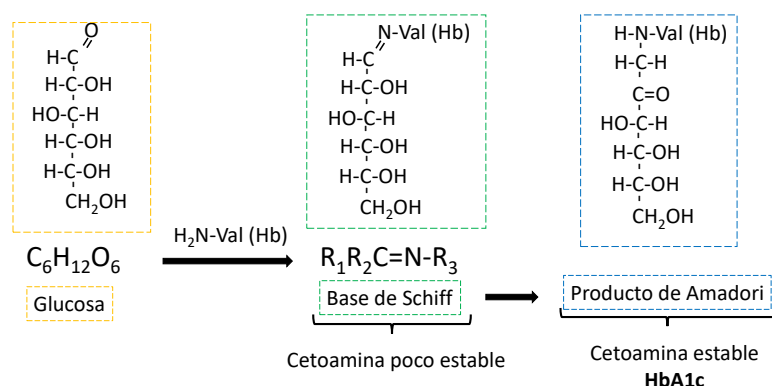
La diabetes es una enfermedad que ocurre cuando el cuerpo no produce y/o no usa la hormona insulina de manera adecuada, generando que se acumule demasiada glucosa (azúcar) en la sangre. Hay 2 tipos principales de diabetes. El tipo 1 ocurre cuando el cuerpo no produce nada de insulina; a veces se llama diabetes juvenil porque generalmente se descubre en niños y adolescentes, pero también puede aparecer en adultos. El tipo 2 ocurre cuando el cuerpo no produce suficiente insulina o no la usa como debería. En el pasado, los médicos pensaban que solo los adultos estaban en riesgo de desarrollar diabetes tipo 2 [14].

La Asociación Estadounidense de Diabetes y otras organizaciones relacionadas en todo el mundo han recomendado la medición de hemoglobina glicosilada (HbA<sub>1c</sub>) para ayudar en el diagnóstico de diabetes; esta prueba consiste en analizar muestras de sangre de pacientes. El porcentaje de HbA<sub>1c</sub> muestra un historial glicémico acumulado del paciente durante los últimos dos o tres meses [15, 16, 17, 18]. El conocer los niveles de glucosa en un paciente diabético es de suma importancia para su control y con ello evitar complicaciones como lo son enfermedades cardiovasculares, nerviosas, enfermedades del riñón, entre otras [19, 20]. Los valores de HbA<sub>1c</sub> asociados al diagnóstico y control de diabetes que se consideran son [21]:

- menor que 5.7 % → No diabéticos
- de 5.7 % a 6.5 % → Pre-diabéticos
- mayor que 6.5 % → Diabéticos

La HbA<sub>1c</sub> se forma mediante la glicosilación no enzimática de la hemoglobina, la cual consiste en el aumento persistente de la glucosa en

la sangre, el cual acelera la reacción entre azúcares y otras moléculas como proteínas, lípidos y ácidos nucleicos produciendo modificaciones por reducción de azúcar. Durante esta reacción se produce la condensación de la glucosa en el grupo N-amino de la cadena  $\beta$  de la hemoglobina A, generando una unión covalente entre el extremo del grupo amino y el carbón carbonílico de la glucosa. Esta unión da lugar a un compuesto poco estable denominado Base de Schiff, Aldimina o HbA<sub>1c</sub> lábil. Posteriormente, se inicia un proceso de reordenamiento de los enlaces químicos, llamado reordenamiento de Amadori, el cual produce un enlace cetamínico estable, que da como resultado al enlace complejo de HbA<sub>1c</sub>, el cual se mide a nivel laboratorio [1, 2, 22]. La figura 1.1 muestra la fórmula de HbA<sub>1c</sub>.



**Figura 1.1:** Fórmula química de la HbA<sub>1c</sub> [1, 2]

La glucosa es otro parámetro importante para el control de los pacientes diabéticos. También llamada dextrosa, la glucosa pertenece a un grupo de carbohidratos conocidos como azúcares simples (monosacáridos); tiene la fórmula molecular  $C_6H_{12}O_6$ , es la fuente de energía en la función celular y la regulación de su metabolismo es de gran importancia [23]. En algunos libros pueden encontrarse diagramas de la molécula de glucosa, luciendo como se observa en la figura 1.1 primera parte, aunque este diagrama de barras se refiere a una distribución en glucosa seca (en polvo); en el cuerpo humano la glucosa generalmente se encuentra de forma de anillo [24]. Los criterios actuales de la Organización Mundial de la Salud (OMS) para diagnosticar diabetes son: glucosa plasmática en ayunas mayor que 7,0 mmol/l (126 mg/dl) o glucosa plasmática medida a las 2h mayor que 11,1 mmol/l (200 mg/dl) [25].

## 1.1. Estado del Arte

La detección de HbA<sub>1c</sub> es de gran importancia; en consecuencia, se han propuesto varios métodos para su análisis y medición, como el inmunoensayo [26, 27, 28], la cromatografía líquida de alta presión [29, 30], la espectrofotometría [31, 32, 33], el intercambio iónico y la cromatografía de afinidad [34, 35, 36, 37], entre otros. La mayoría de estos métodos se basan en la ruptura química de las fracciones de hemoglobina, tanto no glicadas como glicadas [38].

En los últimos años, se han reportado varios enfoques ópticos para la caracterización, identificación y cuantificación de HbA<sub>1c</sub> [39, 40]. Por ejemplo, el trabajo publicado por M. Mallya et al. [41] utilizaron diferentes concentraciones de glóbulos rojos humanos liofilizados y aplicó una regresión lineal apoyada con el área bajo la curva de espectros, para la estimación de los porcentajes de HbA<sub>1c</sub>. En el trabajo de O. S. Zhernovaya et al. [42], dieron a conocer cálculos del índice de refracción en diferentes concentraciones de Hb-glucosa, utilizando tomografía de coherencia óptica (OCT, por sus siglas en inglés), así como refractometría de Abbe [33]. Por otro lado, G. Mazarevica et al. [43], discriminaron Hb de HbA<sub>1c</sub> utilizando el índice de refracción y la tomografía de difracción. El trabajo de E. N. Lazareva et al. [38], presenta una comparación de las propiedades ópticas de HbA<sub>1c</sub> y Hb usando refractometría, fluorescencia y espectroscopía Raman de mejora de superficie (SERS, por sus siglas en inglés). Una revisión de O. A. Smolyanskaya et al. [44], presenta varios estudios espectroscópicos de la glicación de tejidos y proteínas celulares, tanto en organismos que muestran diabetes desarrollada de forma natural como en muestras glicadas *in vitro*, utilizando una amplia gama de ondas electromagnéticas que incluyen refractometría óptica, microscopía de holográfica digital, tomografía de difracción, fluorescencia, espectroscopía de terahercios e imágenes ópticas.

Además, el trabajo de T. Pan et al., [45] presenta un modelo para la discriminación entre pacientes diabéticos y no diabéticos, utilizando glóbulos rojos y espectroscopia Raman de infrarrojo cercano (NIRS, por sus siglas en inglés). E. N. Lazareva et al. [38] utilizaron SERS para analizar el comportamiento espectral del polvo liofilizado de hemoglobina humana (Hb)

y HbA<sub>1c</sub> hemolizada. En el trabajo de I. Barman et al. [46] se reportó un procedimiento para la discriminación de Hb y HbA<sub>1c</sub> tanto en muestras de sangre como comerciales, utilizando espectroscopia Raman de recubrimiento de gotas (DCRS, por sus siglas en inglés); allí se aplicó el método de mínimos cuadrados parciales (PLS, por sus siglas en inglés) para el cálculo del porcentaje de HbA<sub>1c</sub>. En el trabajo de S. Kiran et al. [47] pudieron caracterizar el comportamiento espectral de Hb y HbA<sub>1c</sub> utilizando espectroscopía Raman de resonancia con superficie mejorada (SERRS, por sus siglas en inglés). El trabajo presentado por J. Lin et al. [48] muestra la viabilidad de discriminar pacientes diabéticos tipo 2 de pacientes no diabéticos, utilizando una combinación de pinzas láser y espectroscopia Raman; estos autores analizaron diferencias en los picos de espectros obtenidos a partir de glóbulos rojos de ambos tipos de pacientes. En el trabajo de R. Pandey et al. [49], las señales Raman de Hb y HbA<sub>1c</sub> se analizaron en eritrocitos hemolizados y sangre completa, utilizando espectroscopia de resonancia Raman; se clasificaron las concentraciones de HbA<sub>1c</sub> en tres grupos utilizando una máquina de soporte vectorial (SVM, por sus siglas en inglés); tal trabajo incluye una asignación molecular de picos característicos.

Con respecto a la medición *in vivo*, el trabajo de J. F. Villa-Manríquez et al. [50] discrimina entre pacientes diabéticos y no diabéticos utilizando espectroscopia Raman y SVM. El uso de espectroscopia de terahercios en piel humana para la detección de glucosa se reportó por O. Cherkasova et al. en [51]. El trabajo de D. L. Felipe et al. [52] evaluó las relaciones de los productos finales de la glicación avanzada en la piel, utilizando fluorescencia intrínseca en piel (SIF, por sus siglas en inglés) y regresión multivariable. El trabajo de S. Kim y J. Lee [53] presenta una patente de un aparato que usa Transformada de Fourier infrarroja Raman y espectrómetros de absorbancia para la detección de glucosa y HbA<sub>1c</sub>. Sin embargo, no se proporciona información sobre el método para calcular tales aproximaciones y no se obtuvo ninguna estimación de error de las mediciones para dicho aparato.

Otro parámetro importante en el control de la diabetes es la medición de glucosa, la cual se realiza comúnmente mediante la extracción de 10  $\mu$ l de sangre sérica o plasmática, a la cual se le aplica el método de glucosa oxidasa, que cataliza la oxidación de glucosa a ácido glucónico. El peróxido

de hidrógeno producido es detectado por un aceptor cromogénico de oxígeno; la intensidad de color obtenida es proporcional a la concentración de glucosa en la muestra; se usa un espectrofotómetro a 505 nm para llevar a cabo las mediciones. Esta prueba tiene la desventaja de que otros compuestos séricos pueden oxidarse, como por ejemplo la bilirrubina, el ácido ascórbico y el ácido úrico, debido a la reacción catalizada por la enzima glucosa oxidasa y por lo tanto interferir con el resultado [54, 55].

Entre otros trabajos relacionados al tema de esta investigación, se encuentra el de L. Silveira et al. [56], quienes presentaron un estudio de 44 muestras de suero sanguíneo mediante espectroscopia Raman y regresión de mínimos cuadrados parciales (PLS), obteniendo un valor de raíz del error cuadrático medio (RMSE, por sus siglas en inglés) de 26.7 mg/dL, calculado usando validación cruzada. el trabajo de R. de Cássia Fernandes Borges et al. [57] utilizaron 44 muestras de suero sanguíneo para la detección de adulteraciones mediante un espectrómetro Raman dispersivo de infrarrojo cercano. En dicho trabajo se propuso un modelo basado en PCA, para la extracción de características espectrales; además se utilizó distancia euclidiana para la discriminación entre grupos normales e hiperglucémicos obteniendo 70 % de clasificación correcta. En el trabajo de J. Xue et al. [58] se desarrolló un método utilizando espectroscopía de infrarrojo cercano (NIR) y espectroscopía Raman, investigando dos tipos de plasma artificial con 100 muestras por tipo; los modelos de calibración se generaron mediante la realización de regresión de mínimos cuadrados parciales obteniendo una (RMSEP por sus siglas en inglés) de 0.156 mg/dL. En los trabajos de D. Rohleder et al. [59, 60], los autores presentaron un estudio de suero medido por espectroscopia Raman y de infrarrojo medio combinado con PLS a 247 voluntarios, en el que obtuvieron un RMSEP de 17.1 mg/dL. El trabajo de A. J. Berger et al. [61] presenta un análisis multi-componente de suero de 69 pacientes utilizando espectroscopia Raman y PLS, obteniendo un RMSEP de 26 mg/dL. El trabajo reportado por Jianan Y. Qu et al. [62], los autores realizan mediciones de suero de 60 voluntarios utilizando espectroscopia Raman y el método PLS, en el que obtuvo un RMSEP de 1.3 mmol/L = 23.42 mg/dL. El trabajo de D. R. Parachalil et al. presentó un estudio de 25 sueros ultrafiltrados, es decir, mediante un procedimiento de centrifugación para eliminar la

fracción de alto peso molecular del suero, facilitando el análisis de proteínas y biomarcadores moleculares; obtuvieron un RMSECV de 1,84 mg/dL y un 98 % de aciertos en la Zona A más B en la cuadrícula de error de Clarke [63]. Por otro lado, S. Giansante et al. informaron de un estudio en suero de 242 voluntarios en el que se midieron y predijeron sus valores de colesterol total, triglicéridos, urea, creatinina, colesterol HDL y glucosa mediante espectroscopia Raman y PLS. Sus resultados proporcionaron un RMSECV de 15,4 mg/dL de glucosa que es de interés en esta investigación [64].

Con respecto a mediciones *in vivo* de glucosa hay varios trabajos disponibles. Por ejemplo, en el trabajo de J. Shao et al. [65], usando ratas con 3 concentraciones que inyectaron en la oreja del animal, analizaron el pico de  $1125\text{ cm}^{-1}$ , y obtuvieron un error absoluto medio (MAE, por sus siglas en inglés) en la validación del 6.7 % y un  $0.95\text{ R}^2$ . En el trabajo de M. J. Scholtes-Timmerman et al. [66] se midieron 111 sujetos utilizando un sistema de espectrómetro Raman hecho a medida, obtuvieron resultados acordes a la rejilla de error de Clarke de 78.4 % en la zona A, 20.7 % en la zona B y 0.9 % en la zona D. Asimismo, S. M. Lundsgaard-Nielsen et al. [67] analizaron 35 pacientes con diabetes durante un período de 60 días, utilizando Raman de profundidad crítica (CD-Raman, por sus siglas en inglés) y PLS; obtuvieron un porcentaje de acierto utilizando la rejilla de Clarke con un 54.1 % en la zona A y B. En el trabajo de K. Ma et al. [68] implementaron un sensor basado en espectroscopía Raman con compensación espacial mejorada en superficie (SESORS, por sus siglas en inglés), el cual implantaron en 5 ratas, obteniendo un RMSEP de 13.7 mg/dL. Asimismo, J. Chaiken et al. [69] realizaron 49 mediciones en voluntarios con diabetes tipo 1 y 2; utilizaron la rejilla de error de Clarke para determinar su precisión con un 72 % en la región A. Los autores Annika M. K. Enejder et al. [70] midieron 17 sujetos humanos sanos cuyos niveles de glucosa en sangre se elevaron durante un período de 2 a 3 hrs. utilizando un protocolo de prueba de tolerancia a la glucosa estándar, obteniendo un error absoluto medio (MAE) de 7.8 % y  $\text{R}^2$  de 0.87. En el trabajo de D. A. Stuart et al. [71] utilizaron espectroscopia Raman se superficie mejorada en un ratas obteniendo un RMSEP de 53.42 mg/dL. De igual manera, J. W. Kang et al. [72] realizaron mediciones en la oreja de 3 cerdos obteniendo

un resultado de 72.8 % de aciertos en la rejilla de error de Clarke. J. Chaiken et al. reportan una revisión bibliográfica del uso de espectroscopia Raman para analizar sangre y tejido para determinar glucosa [73]. Cabe mencionar que se han reportado otros trabajos recientes, sin embargo estos son usando diferentes métodos espectroscópicos como por ejemplo espectroscopia Raman de transformada de Fourier [74], usando espectroscopia Raman de superficie mejorada (SERS) para determinar glucosa *in vitro* [75], así como usando espectroscopia de reflectancia en infrarrojo cercano [76], de igual manera utilizando espectroscopia Raman y comprendiendo la dinámica conocida de la matriz de tejido sanguíneo [77], entre otros. Sin embargo, esta bibliografía fue lo encontrado hasta la fecha de cierre del trabajo (mayo 2021), en donde solamente espectroscopia Raman de emisión espontánea fue utilizada.

## 1.2. Hipótesis

Sabiendo que mediante la interacción de la luz con determinado tejido biológico y/o biomarcador se producen vibraciones de las moléculas que lo componen es posible utilizar modelos basados en espectroscopia Raman y Redes Neuronales Artificiales para aproximar cuantitativamente el porcentaje de HbA<sub>1c</sub> y la cantidad de glucosa *in vivo* en pacientes sanos y diabéticos.

## 1.3. Objetivos

### 1.3.1. Objetivo general

Diseñar un proceso sistemático y congruente, capaz de cuantificar porcentajes de glucosa y HbA<sub>1c</sub> en soluciones e *in vivo*, además de otras sustancias relacionadas, mediante el uso de espectroscopia Raman y regresiones basadas en ANN.

### 1.3.2. Objetivos particulares

- Diseñar modelos de ANN que, calibrados en base a ejemplos, obtengan los mejores resultados posibles en la determinación de mediciones de

glucosa y HbA<sub>1c</sub> *in vivo* en soluciones a diferentes concentraciones, así como en mediciones obtenidas *in vivo*.

- Probar los modelos generados en el punto anterior en diferentes concentraciones de sacarosa y fructosa.
- Adaptar un prototipo diseñado en el Grupo de Instrumentación y Óptica Biomédica (GIOB) de la coordinación de Óptica del INAOE para obtener mediciones de espectros Raman en sujetos humanos *in vivo*.
- Diseñar un protocolo de adquisición y una base de datos de mediciones de espectros Raman obtenidas de sujetos humanos.
- Analizar el desempeño de los modelos diseñados en comparación a otros modelos de estimación populares, tales como PCA y SVM.
- Identificar los picos característicos en los espectros Raman obtenidos en sustancias puras y compararlos con los en espectros *in vivo*.

## 1.4. Resumen del documento

Este documento se presenta de la siguiente manera: el capítulo dos introduce los conceptos principales, los cuales abarcan la técnica óptica y los métodos computacionales aplicados para el mejoramiento de la señal y predicción de concentraciones de glucosa y HbA<sub>1c</sub>. El capítulo tres muestra el desarrollo experimental, incluyendo las mediciones *in vivo* e *in vitro*, así como las mediciones de sustancias puras. El capítulo cuatro presenta los procesos de medición, es decir, los métodos computacionales utilizados para obtener la concentración de glucosa y HbA<sub>1c</sub>. El capítulo cinco muestra los resultados del proceso de aproximación de los valores de las sustancias analizadas. En el capítulo seis se presentan las conclusiones, las contribuciones de este trabajo al estado del arte y se discute sobre posibles mejoramientos de esta investigación; así mismo se mencionan los artículos y participaciones en conferencias derivadas de la investigación.



---

## MARCO TEÓRICO

---

En este capítulo se abordarán los conceptos principales en los que se basó el desarrollo de la presente tesis. Dichos conceptos incluyen la técnica de adquisición de datos, la cual está basada en espectroscopia Raman, así como los métodos computacionales utilizados para la cuantificación de las concentraciones de las sustancias analizadas, tales como redes neuronales artificiales, máquina de soporte vectorial, regresión lineal, regresión de mínimos cuadrados por intervalos, así como técnicas de extracción y selección de características.

### 2.1. Espectroscopia Raman

La espectroscopia Raman es una técnica de espectroscopia vibracional, que se puede utilizar para analizar ópticamente los cambios moleculares asociados con la muestra en estudio [78]. Las bandas espectrales en los espectros vibracionales son específicas de una molécula y proporcionan información directa sobre la composición bioquímica [79].

#### 2.1.1. Efecto Raman

Las moléculas tienen la capacidad de esparcir luz. En este proceso, la dirección del fotón saliente cambia con respecto al incidente, y en la mayoría de veces la longitud de onda del fotón permanece constante. Esto quiere decir que se presenta una interacción elástica conocida como esparcimiento Rayleigh, el cual se usa ampliamente para medir el tamaño de las partículas y sus distribuciones, hasta tamaños inferiores a  $1\text{ }\mu\text{m}$  [80]. Sin embargo, en 1923, Adolf Smekal predijo la existencia de una posibilidad aún más rara de

esparcimiento inelástico. Este efecto fue luego descubierto experimentalmente por Raman y Krishnan en 1928, por lo que más tarde recibió el nombre de su principal descubridor como el "efecto Raman" [81].

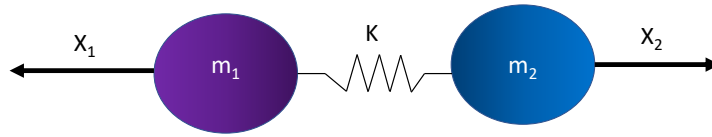
Raman y su grupo de trabajo realizaron una serie de mediciones de la luz esparcida por líquidos y sólidos; en uno de sus experimentos la luz violeta del espectro solar fue aislada mediante un filtro violeta y paso hacia una muestra líquida. La mayoría de la luz que emergió de la muestra líquida fue del mismo color que el haz violeta incidente (esparcimiento Rayleigh de la luz). Sin embargo, Raman y K.S Krishnan fueron capaces de demostrar que parte de la luz esparcida era de diferente color, la cual pudieron aislar usando un filtro verde, colocado entre el observador y la muestra, como se muestra en la figura 2.1 [3, 82].



**Figura 2.1:** Experimento de Chandrasekhara Venkata Raman [3].

Al considerar el esparcimiento Raman, se puede pensar dos maneras: la interpretación de onda clásica o la interpretación de partículas cuánticas. En la interpretación de ondas clásica, la luz se considera como radiación electromagnética, que contiene un campo eléctrico oscilante interactuando con una molécula a través de su polarización. En general, es más fácil iniciar con la interpretación clásica considerando una molécula diatómica, como una masa en un resorte, como se observa en la figura 2.2, donde  $m$  representa la masa atómica,  $x$  representa el desplazamiento y  $K$  representa la fuerza de unión [83, 80, 84].

Cuando se usa esta aproximación, el desplazamiento de la molécula se



**Figura 2.2:** Molécula diatómica con una masa en un resorte [4].

puede expresar usando la ley de Hooke como:

$$\frac{m_1 m_2}{m_1 + m_2} \left( \frac{d^2 x_1}{dt^2} + \frac{d^2 x_2}{dt^2} \right) = -K(x_1 + x_2) \quad (2.1)$$

Remplazando la masa reducida ( $m_1 m_2 / [m_1 + m_2]$ ) con  $\mu$  y el desplazamiento total ( $x_1 + x_2$ ) con  $q$ , la ecuación se puede simplificar a:

$$\mu \frac{d^2 q}{dt^2} = -Kq \quad (2.2)$$

Resolviendo esta ecuación para el desplazamiento  $q$  tenemos:

$$q = q_0 \cos(2\pi \nu_m t) \quad (2.3)$$

Donde  $\nu_m$  es la vibración molecular definida como:

$$\nu_m = \frac{1}{2\pi} \sqrt{\frac{K}{\mu}} \quad (2.4)$$

De la ecuación 2.3 y 2.4 se puede observar que la molécula vibra siguiendo un patrón coseno con una frecuencia proporcional a la fuerza del enlace e inversamente proporcional a la masa reducida. De esto podemos ver que cada molécula tendrá sus propias firmas vibratorias únicas que están determinadas no solo por los átomos de la molécula, sino también por las características de los enlaces individuales [4, 80, 84, 79].

Cuando la luz incidente interactúa con una molécula, induce un momento dipolar  $P$ , igual al producto de la polarizabilidad de la molécula y el campo eléctrico de la fuente de luz incidente [4, 80, 84]. Esto se puede expresar como:

$$P = \alpha E_0 \cos(2\pi \nu_0 t) \quad (2.5)$$

Donde  $E_0$  es la intensidad y  $\nu_0$  es la frecuencia del campo eléctrico. Utilizando la aproximación de pequeña amplitud, la polarizabilidad puede describirse como una función lineal del desplazamiento:

$$\alpha = \alpha_0 + q\left(\frac{\partial\alpha}{\partial t}\right)_{q=0} + \dots \quad (2.6)$$

Cuando se combina con las ecuaciones 2.3 y 2.5 da como resultado:

$$P = \alpha_0 E_0 \cos(2\pi\nu_0 t) + q_0 \cos(2\pi\nu_m t) E_0 \cos(2\pi\nu_0 t) \left(\frac{\partial\alpha}{\partial t}\right)_{q=0} \quad (2.7)$$

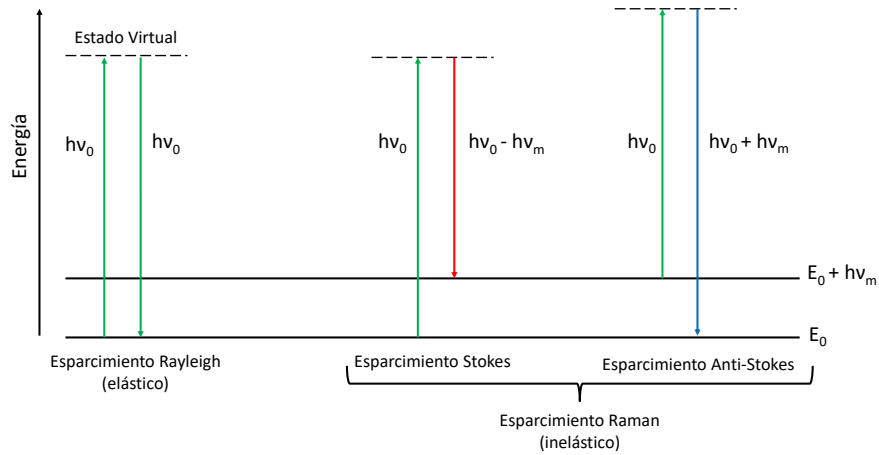
En la ecuación 2.7 vemos que hay dos efectos resultantes de la interacción de la molécula y la luz incidente. El primer efecto se llama esparcimiento de Rayleigh, que es el efecto dominante y no produce cambios en la frecuencia de la luz incidente [84]. El segundo efecto es el componente de esparcimiento Raman y cuando se expande a:

$$q_0 E_0 \left(\frac{\partial\alpha}{\partial t}\right)_{q=0} [\cos(2\pi\{\nu_0 - \nu_m\}t) + \cos(2\pi\{\nu_0 + \nu_m\}t)] \quad (2.8)$$

Se puede demostrar que cambia la frecuencia de la luz incidente en más o menos la frecuencia de la vibración molecular. El aumento de frecuencia se conoce como un cambio de Anti-Stokes y la disminución de la frecuencia se conoce como un cambio de Stokes. Al medir el cambio en la frecuencia de la luz incidente (normalmente solo se usa el cambio de Stokes para esta medición), el efecto Raman ahora brinda a los espectroscopistas un medio para medir directamente la frecuencia vibratoria de un enlace molecular [4, 84]. Ahora que se ha obtenido la interpretación del efecto Raman a partir de modelo de onda clásico, se puede usar la interpretación cuántica para determinar información adicional. Como se presentó anteriormente, el efecto Raman se describe como el esparcimiento inelástico de un fotón fuera de un enlace molecular. Una manera de observar el resultado del fotón incidente que excita a la molécula a un estado de energía virtual, es mediante el diagrama de Jablonski que se muestra en la figura 2.3 [5].

Cuando esto ocurre, hay tres posibles resultados diferentes [84]:

- La molécula puede volver a relajarse hasta el estado fundamental y emitir un fotón de igual energía que el fotón incidente; se trata de un proceso elástico y se denomina esparcimiento de Rayleigh.



**Figura 2.3:** Diagrama de Jablonski que representa transiciones de energía cuántica para el espárcimiento de Rayleigh y Raman [5, 6].

- Si el fotón espárcido tiene una frecuencia menor a la del incidente, se produce una transferencia de energía del fotón a la molécula que, después de saltar al estado de energía no permitido, vuelve a uno permitido mayor al que tenía inicialmente ( $\nu_0 - \nu_r$ ); esto se denomina espárcimiento Raman Stokes.
- Si la molécula ya se encuentra en un estado virtual de mayor energía y luego se relaja de nuevo al estado fundamental emitiendo un fotón con una frecuencia mayor a la incidente ( $\nu_0 + \nu_r$ ); esto se llama espárcimiento Raman Anti-Stokes [6].

Debido al hecho de que la mayoría de las moléculas se encontrarán en el estado fundamental a temperatura ambiente, existe una probabilidad mucho menor de que un fotón presente espárcimiento Anti-Stokes. Como resultado, la mayoría de las mediciones Raman se realizan considerando solo la luz desplazada de Stokes. Al investigar más a fondo la interpretación cuántica del efecto Raman, se puede demostrar que la potencia de la luz espárcida  $\bar{P}_s$ , es igual al producto de la intensidad de los fotones incidentes  $\bar{I}_0$ , y un valor conocido como sección transversal Raman,  $\sigma_R$  [85]. Se puede demostrar que:

$$\sigma_R \propto \frac{I_0}{\lambda^4} \quad (2.9)$$

Donde  $\lambda$  equivale a la longitud de onda del fotón incidente, por lo tanto:

$$P_s \propto \frac{I_0}{\lambda^4} \quad (2.10)$$

De la ecuación 2.10 se desprende que existe una relación lineal entre la potencia de la luz esparcida y la intensidad de la luz incidente, así como una relación entre la potencia de la luz esparcida y el inverso de la longitud de onda a la cuarta potencia. Por tanto, parecería que siempre es deseable utilizar una longitud de onda de excitación corta y una fuente de excitación de alta potencia basada en estas relaciones. Sin embargo, no siempre es posible pues depende de la muestra a analizar [85, 4].

La Tabla 2.1 presenta algunas vibraciones moleculares típicas con sus respectivos desplazamientos Raman, donde  $\tilde{\nu}$  corresponde al número de onda [86, 87].

### 2.1.2. Tipos de ruido en espectroscopia Raman

Los espectros Raman se conforman por señales bidimensionales, que, como muchos otros tipos de señales, contienen ruido. Definiremos el ruido como el conjunto de señales que no son de interés o que ocasionan perturbaciones al analizar un conjunto de datos. Para las mediciones de espectros Raman, los ruidos más comunes se pueden clasificar en cinco grupos diferentes: ruido de disparo, ruido generado por la muestra, ruido generado por instrumentación, ruido de computadora y ruido generado por fuentes externas [88, 89].

**Ruido de disparo:** consiste en fluctuaciones aleatorias de la corriente eléctrica del dispositivo a través de un conductor, provocadas por el hecho de que la corriente se realiza en cargas discretas (electrones) [90].

**Ruido generado por la muestra:** incluye emisiones ópticas no deseadas, como la fluorescencia, que generalmente se presenta como una curvatura suave de la línea base del espectro. Puede alcanzar intensidades que enmascaran por completo la intensidad de las bandas Raman. Además, incluye la variación de la intensidad de la señal debido a cambios de temperatura o heterogeneidad en la composición del muestreo [89].

**Ruido generado por instrumentación:** incluye alteraciones introducidas por el detector como por ejemplo el ruido térmico o de Johnson, que se

**Tabla 2.1:** Vibraciones típicas de algunas moléculas y enlaces.

| Tipo de vibración             | $\tilde{\nu}/\text{cm}^{-1}$ |
|-------------------------------|------------------------------|
| C-H estiramiento              | 2850 - 2960 $\text{cm}^{-1}$ |
| C - H doblez                  | 1340 - 1465 $\text{cm}^{-1}$ |
| C-C estiramiento, doblez      | 700 - 1250 $\text{cm}^{-1}$  |
| C = C estiramiento            | 1620 - 1680 $\text{cm}^{-1}$ |
| C C $\equiv$ estiramiento     | 2100 - 2260 $\text{cm}^{-1}$ |
| O - H estiramiento            | 3590 - 3650 $\text{cm}^{-1}$ |
| H- enlaces                    | 3200 - 3570 $\text{cm}^{-1}$ |
| C = O estiramiento            | 1640 - 1780 $\text{cm}^{-1}$ |
| C $\equiv$ N estiramiento     | 2215 - 2275 $\text{cm}^{-1}$ |
| N - H estiramiento            | 3200 - 3500 $\text{cm}^{-1}$ |
| C- F estiramiento             | 1000-1400 $\text{cm}^{-1}$   |
| C - Cl estiramiento           | 600 - 800 $\text{cm}^{-1}$   |
| C - Br estiramiento           | 500 - 600 $\text{cm}^{-1}$   |
| C-I estiramiento              | 500 $\text{cm}^{-1}$         |
| CO <sub>2</sub> <sup>-3</sup> | 1410 - 1450 $\text{cm}^{-1}$ |
| NO <sub>3</sub> <sup>3</sup>  | 1350 - 1420 $\text{cm}^{-1}$ |
| NO <sub>2</sub> <sup>2</sup>  | 230 - 1250 $\text{cm}^{-1}$  |
| Silicatos                     | 900 - 1100 $\text{cm}^{-1}$  |

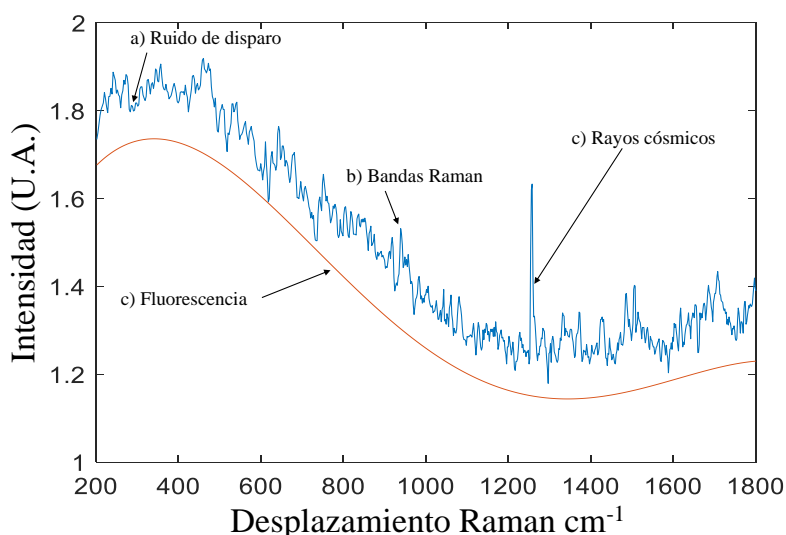
produce por la agitación térmica de los portadores de corriente en un elemento resistivo. Otro es el ruido de temperatura, debido a los procesos estadísticos de intercambio de calor entre un detector y su entorno, que produce fluctuaciones de la señal eléctrica. Un tercer tipo es el ruido de recombinación de generación, debido a las estadísticas de los portadores de carga. El ruido de contacto se genera por la fluctuación de corriente a través de los contactos eléctricos; el ruido de radiación se produce debido a fluctuaciones estadísticas de los flujos de fotones según las estadísticas de Poisson; el ruido de corriente oscura es debido a la suma de las corrientes de ruido en ausencia de señal [91].

**Ruido computacional:** se refiere al ruido introducido durante el proceso de escaneo de la señal de salida del detector.

**Ruido generado por fuentes externas:** incluye el ruido generado por partículas de alta energía, como los rayos cósmicos, que dan como resultado

picos de esparcimiento muy estrechos e intensos en el espectro Raman; los picos pueden ocurrir en momentos y posiciones aleatorias [92].

La figura 2.4 muestra algunos de estos ruidos en un espectro Raman de muestra biológica.



**Figura 2.4:** Ejemplo de ruidos en el espectros Raman: a) Ruido de disparo, b) Bandas Raman, c) Ruido generado por la muestra (fluorescencia), d) Ruido generado por fuentes externas (rayos cósmicos).

### 2.1.3. Eliminación de fluorescencia y suavizado

La eliminación de la fluorescencia y ruidos en la espectroscopia Raman ha sido suficientemente estudiada; se han propuesto diferentes métodos con esta finalidad, tales como: Ajuste de curvas polinomial [93], corrección de la línea base (baseline, en inglés) [88], descomposición de modo empírico [94], el algoritmo Vancouver Raman [94], Wavelets [95], estimación de máxima verosimilitud [96] y Savitzky-Golay [97]. Para los experimentos reportados en esta investigación, se utilizaron dos propuestas estudiadas en el Grupo de Instrumentación y Óptica Biomédica (GIOB). La primera busca eliminar el ruido de disparo y suavizar la señal mediante el uso de *wavelets*; la otra propuesta busca la eliminación de fluorescencia mediante ajuste de los polinomios de Zernike usando algoritmos genéticos. A continuación, se describen brevemente



ambas técnicas; para más detalles [95, 98, 99].

- **Wavelets:** se usan como funciones base para representar otras funciones. En el caso de la espectroscopia Raman, el espectro resultante puede considerarse como la señal a transformar  $f(x)$ , donde el eje  $x$  corresponde al desplazamiento Raman, puesto que la señal presenta dependencia de la intensidad (luz esparcida inelásticamente) en desplazamiento Raman  $\Delta(1/\lambda)$  de los fotones después de su interacción con la materia. El algoritmo de wavelets divide la señal  $f(x)$  en diferentes componentes de frecuencias. En resumen, la señal  $f(x)$  es representada en términos de la suma de wavelets elementales, para más información de los parámetros utilizados consultar [99].
- **Polinomios de Zernike y algoritmos genéticos:** Los polinomios de Zernike sirven para resolver problemas ópticos de interferencia; en este caso se utilizan para ajustar la curva de fluorescencia. Los algoritmos genéticos se utilizan como una técnica de optimización y búsqueda; una de sus aplicaciones es ajustar curvas y funciones a partir de datos. Similar a otros algoritmos de optimización, los algoritmos genéticos buscan un mínimo global donde converge la solución. Una vez identificando el mejor ajuste de curva, ésta se resta al espectro original, con lo cual se consigue la eliminación de fluorescencia, para más información de la implementación y parámetros utilizados consultar [98].

## 2.2. Aprendizaje de máquina

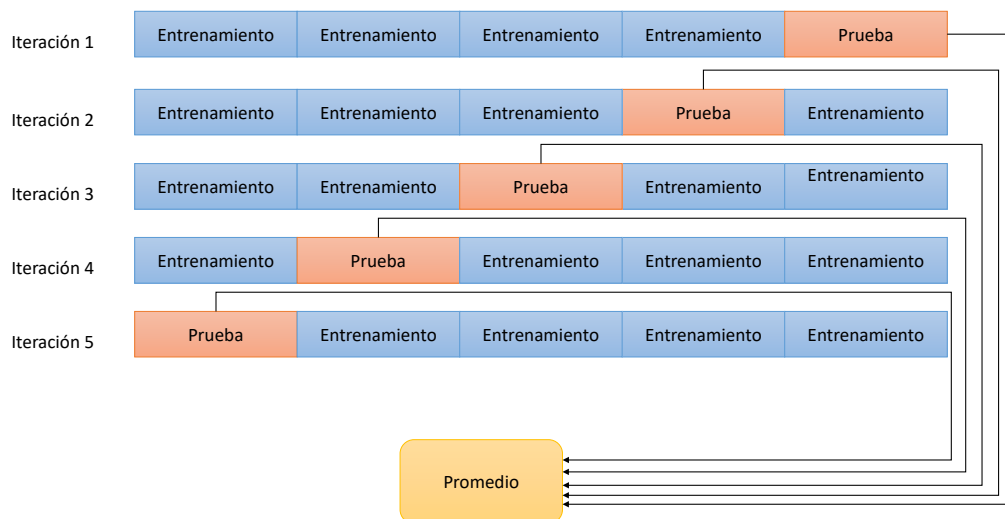
El aprendizaje de máquina ( *Machine Learning*, en inglés) es una sub-área de las ciencias computacionales en la que se desarrollan técnicas con las cuales los dispositivos programables pueden “aprender” sobre determinada problemática, a partir de información disponible. En términos generales, aprendizaje es el proceso de convertir experiencia en conocimiento. Las entradas a un algoritmo de aprendizaje se conocen como datos de entrenamiento, lo cuales representan la experiencia; el resultado del aprendizaje es la representación numérica de cierto conocimiento adquirido a partir de los datos de entrada. Este aprendizaje generalmente toma la forma de otro programa de computadora que puede

realizar alguna tarea [100]. En el aprendizaje de máquina se usan diversos tipos de entrenamiento en función de las características del problema, los cuales pueden ser:

1) Entrenamiento supervisado, el cual requiere que se conozcan, además del conjunto de datos correspondientes a las entradas que representan el problema, los respectivos valores de salida esperada por el sistema.

2) Entrenamiento no supervisado, el cual implica que lo único que ingresa al algoritmo son las entradas del sistema; las características que presentan las entradas se agrupan de forma auto-organizada a través de técnicas de reforzamiento [7].

Considerando que el éxito de estos sistemas depende fuertemente de los datos utilizados para entrenarlos, un punto importante en el aprendizaje de máquina es la manera de dividir los datos utilizados para su entrenamiento y evaluación. Una forma de evaluación es la cruzada, en donde los datos se dividen en  $n$  partes, como se muestra en la figura 2.5. Cada iteración considera el conjunto de prueba como un conjunto diferente entre las  $n$  divisiones; el desempeño del sistema se calcula como el promedio del desempeño de éste en cada iteración. De esta manera, también se evita el sobre ajuste en los parámetros del modelo de aprendizaje [101, 102].



**Figura 2.5:** Representación gráfica de una validación cruzada de 5 particiones.

A continuación, se describen las técnicas de aprendizaje de máquinas

utilizadas en esta investigación [103, 50, 104, 105].

### 2.2.1. Redes Neuronales Artificiales

Las redes neuronales artificiales (ANN, por sus siglas en inglés) son modelos matemáticos utilizados para estimar las salidas de una función descrita por un conjunto de entradas. Los elementos fundamentales de las ANN se llaman neuronas, las cuales pueden estar organizadas de varias formas, siendo la organización por capas una de las más populares [106]. Cada neurona puede relacionarse con otras a través de una conexión, que tiene un valor asociado llamado peso. Cuando se utiliza entrenamiento supervisado, los pesos se ajustan utilizando ejemplos conocidos de las entradas y salidas de la función a estimar. La neurona es responsable de calcular la suma ponderada de sus entradas, y éstas son filtradas por una función conocida como función de activación [7].

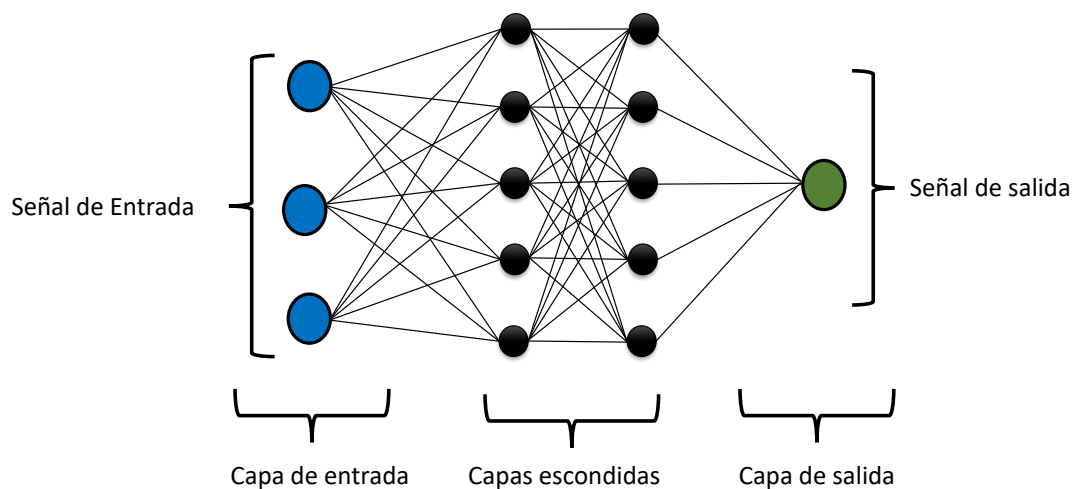
El perceptrón ocupa un lugar especial en el desarrollo histórico de las ANN; fue la primera red neuronal descrita algorítmicamente; éste fue construido alrededor de una sola neurona y se limita a realizar una clasificación de patrones con solo dos clases. Sin embargo, las clases tienen que ser linealmente separables para que el perceptrón funcione correctamente [7]. A partir de esto se generaron diversas arquitecturas, siendo entre las más populares la de perceptrones multicapa o redes alimentadas hacia adelante (FFNN, por sus siglas en inglés). La estructura básica de una FFNN se presenta en la figura 2.6. Las FFNN pueden utilizarse para la solución de problemas de clasificación, regresión y agrupamiento (*Clustering*).

La estructura básica de una neurona se observa en la figura 2.7, donde  $x_1, x_2, \dots, x_n$  son las señales de entrada a la neurona,  $w_{k1}, w_{k2}, \dots, w_{kn}$  son los pesos sinápticos de la neurona  $k$ , y  $y_k$  representa su salida.

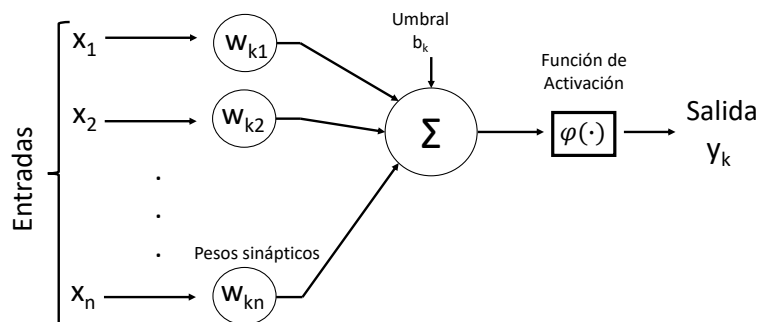
La salida  $y_k$  de una neurona se calcula como:

$$y_k = \varphi \left( \sum_{j=1}^n x_j w_{kj} - b_k \right) \quad (2.11)$$

Siendo  $\varphi$  la función de activación y  $b_k$  el umbral (*bias*, en inglés). El objetivo de la función de activación es unir las salidas de una neurona y suavizar el resultado. Existen diferentes tipos de funciones de activación que pueden



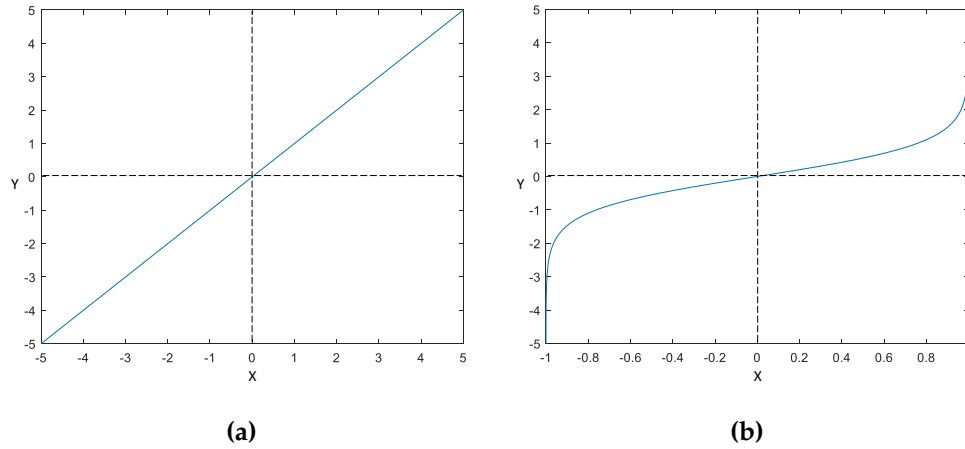
**Figura 2.6:** Gráfico arquitectónico de un perceptrón multicapa o FFNN con dos capas ocultas [7].



**Figura 2.7:** Estructura básica de la unidad fundamental de una red neuronal [7].

ser lineales o no lineales; algunas se muestran en la figura 2.8. Las funciones lineales se utilizan en arquitecturas de redes neuronales simples, o en la última capa de una FFNN cuando se utiliza como un aproximador de función [107].

Desde el punto de vista de la optimización, el aprendizaje de una red equivale a minimizar una función de error global, es decir, que mediante la variación iterativa de los pesos en la ANN se reduzca el error producido entre el cálculo de salida de la ANN con respecto al valor esperado, para cada una de las entradas disponibles en el conjunto de entrenamiento. El aprendizaje se realiza mediante algoritmos de entrenamiento, como por ejemplo el gradiente conjugado escalado y el algoritmo de Levenberg-Marquardt; véase [108, 109] para más información.



**Figura 2.8:** Funciones de activación a) Lineal, b) Sigmoidea tangente hiperbólica [8, 9].

La regresión es el análisis que permite comprender como varía el valor de una variable dependiente con respecto al valor de la variable independiente [110]. Las ANN pueden aproximar funciones con precisión a partir de cualquier conjunto de datos que representen incluso relaciones no lineales [7].

Para los experimentos reportados en esta investigación se utilizó como medida de rendimiento a la raíz del error cuadrático medio (RMSE por sus siglas en inglés); su cálculo se presenta en la ecuación 2.12.  $R_v$  corresponde al valor real esperado de la aproximación,  $P_v$  es el valor predicho obtenido por el modelo siendo evaluado y  $n$  corresponde al número total de muestras [111].

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (R_v - P_v)^2}{n}} \quad (2.12)$$

#### 2.2.1.1. Clasificación

El reconocimiento de patrones es la ciencia de tomar datos sin procesar y realizar una acción decisiva basada en la categoría de un patrón. Un patrón es una entidad etiquetada, representada por un conjunto de propiedades medibles asignadas a una clase definida. Desde un punto de vista matemático, la clasificación consiste en dividir un espacio de  $n$  dimensiones en  $c$  particiones, donde el espacio representa todos los valores posibles de  $n$  características

relevantes que definen un patrón; cada región corresponde a una de las  $c$  clases posibles [112].

El desempeño de un sistema de clasificación puede analizarse de varias formas. Entre las más populares se encuentran las matrices de confusión y la precisión. Una matriz de confusión es cuadrada y contiene tantas filas y columnas como clases en el problema. Cada celda  $(i, j)$  contiene el número de muestras pertenecientes a la clase  $j$  que se clasificaron como pertenecientes a la clase  $i$ . Por ejemplo, en la figura 2.9 se muestra una matriz de confusión para 3 clases. La celda  $TP_A$  (verde) contiene el número de aciertos de la clase “A”,  $TP_B$  el correspondiente a “B” y  $TP_C$  para “C”. La celda  $E_{AB}$  muestra los casos de la clase “A” que fueron clasificados erróneamente como clase “B” y de forma similar ocurre para  $E_{AC}$ ,  $E_{BA}$ , etc.

|   | A        | B        | C        |
|---|----------|----------|----------|
| A | $TP_A$   | $E_{BA}$ | $E_{CA}$ |
| B | $E_{AB}$ | $TP_B$   | $E_{CB}$ |
| C | $E_{AC}$ | $E_{BC}$ | $TP_C$   |

**Figura 2.9:** Matriz de confusión 3 clases [10].

Por otra parte, la precisión se define como [113]:

$$\text{Precisión} = \frac{\text{Predicciones correctas}}{\text{Total de predicciones}} * 100 \quad (2.13)$$

Otra manera de medir la precisión es mediante la métrica sensibilidad o tasa de verdaderos positivos “TPR” y la métrica especificidad o tasa de verdaderos negativos “TNR” [10].

Para problemas de clasificación binaria la sensibilidad se calcula como:

$$TPR = \frac{TP}{TP + FN} \quad (2.14)$$

donde  $TP$  representa el total de verdaderos positivos, es decir, el número de veces que se esperaba una clase positiva y realmente el clasificador obtuvo ésta.

La variable  $FN$  representa los falsos negativos, es decir, el número de veces que se esperaba una clase positiva pero el clasificador obtuvo una clase negativa.

La especificidad se obtiene como:

$$TNR = \frac{TN}{FP + TN} \quad (2.15)$$

En donde  $TN$  es el total de verdaderos negativos por clase, es decir, el número de veces que se esperaba un valor negativo y realmente se obtuvo. El valor  $FP$  representa los falsos positivos, es decir, el número de veces que se obtuvo un valor positivo cuando se esperaba una clase negativa.

Para problemas multi-clase, la especificidad y sensibilidad pueden calcularse por cada clase presente, utilizando la matriz de confusión.

Por ejemplo, en un problema de 3 clases (A, B, C),  $FN$  para la clase "A" se calcula como:

$$FN_A = E_{AB} + E_{AC} \quad (2.16)$$

De forma similar,  $FP$  para la clase "A" se calcula como:

$$FP_A = E_{BA} + E_{CA} \quad (2.17)$$

Para determinar  $TP$  en la clase "A" basta con visualizar la matriz de confusión, los cuadros de color verde corresponden a los  $TP$  de cada clase (véase figura 2.9 parte superior derecha), en el caso de  $TN$  para la misma clase se calcula como:

$$TN_A = TP_B + TP_C \quad (2.18)$$

### 2.2.2. Máquina de Soporte Vectorial

La Máquina de Soporte Vectorial (SVM por sus siglas en inglés) es un algoritmo de aprendizaje supervisado, capaz de realizar una clasificación discriminativa, prediciendo datos no vistos antes; asimismo este método es ocupado para realizar aproximación de datos usando regresión [114]. La SVM mapea los puntos de entrada a un espacio de características de mayor dimensión, luego encuentra el hiperplano que los separa y maximiza el margen entre clases [115], lo cual implica que se debe construir un hiperplano que separe las clases o valores. La solución de hiperplano óptima debe describirse

como la combinación de unos pocos puntos de entrada que se denominan vectores de soporte. Sin ningún conocimiento del mapeo, la SVM encuentra el hiperplano óptimo usando el producto punto con funciones en el espacio de características llamadas Kernels o núcleos [116].

Las funciones del núcleo pueden ser de varios tipos, por ejemplo, los núcleos polinomiales lineales, cuadráticos o cúbicos y los núcleos gaussianos finos, medios o centrales. La ecuación 2.19 presenta un kernel polinomial de orden  $d$ ; la ecuación (2.20) muestra un kernel gaussiano con ancho de núcleo  $\sigma$  [117]. Las variables  $x_i$  y  $x_j$  son vectores en el espacio de entrada (características para el entrenamiento).

$$k(x_i, x_j) = (1 + x_i \cdot x_j)^d \quad (2.19)$$

$$k(x_i, x_j) = \exp \left\{ -\frac{\|x_i - x_j\|^2}{2\sigma^2} \right\} \quad (2.20)$$

### 2.2.3. Regresión lineal

Este análisis ayuda a comprender como varia el valor de la variable dependiente con el valor de la variable independiente [110]. La regresión lineal (LR, por sus siglas en inglés) es probablemente el enfoque más simple para el aprendizaje estadístico. Es un buen punto de partida para enfoques más avanzados y, de hecho, muchas técnicas sofisticadas de aprendizaje estadístico pueden verse como una extensión de la regresión lineal [118, 119]. Por lo tanto, comprender este modelo simple construirá una buena base antes de pasar a enfoques más complejos. La regresión lineal simple se expresa como:

$$y = \beta_0 + \beta_1 X \quad (2.21)$$

En la ecuación anterior,  $\beta_0$  es la ordenada en el origen y  $\beta_1$  es la pendiente de la recta; estos coeficientes son necesarios para hacer predicciones con el modelo. Para encontrar los parámetros, se necesita minimizar la suma de los errores cuadrados. Por supuesto, el modelo lineal no es perfecto y no será capaz de predecir todos los datos con precisión, lo que significa que hay una diferencia entre el valor real y la predicción [118, 119]. La predicción puede estar por encima o por debajo del valor real, lo que da como resultado una



diferencia negativa o positiva respectivamente, por lo cual se eleva el error al cuadrado. Si no se elevan al cuadrado los errores, la suma de los errores podría disminuir debido a diferencias negativas y no porque el modelo sea un buen ajuste. Además, elevar al cuadrado los errores penaliza las grandes diferencias, por lo que minimizar los errores cuadráticos “garantiza” un mejor modelo. Los coeficientes para calcular la recta de regresión de  $y$  sobre  $x$  se calculan de la siguiente manera [118, 119]:

$$\beta_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (2.22)$$

$$\beta_0 = \bar{y} - \beta_1 \bar{x} \quad (2.23)$$

#### 2.2.4. Regresión de mínimos cuadrados parciales por intervalos

La regresión de mínimos cuadrados parciales por intervalo (iPLS, por sus siglas en inglés) propuesto por L. Nøgaard y colaboradores, presenta un modelo en el que la Regresión de Mínimos Cuadrados Parciales (*Partial Least Square PLS*, en inglés) se calcula en subintervalos equidistantes de la región del espectro completo. Esto proporciona una imagen general de la información relevante en diferentes subdivisiones espectrales, centrándose así en regiones espectrales importantes y eliminando interferencias de otras regiones.

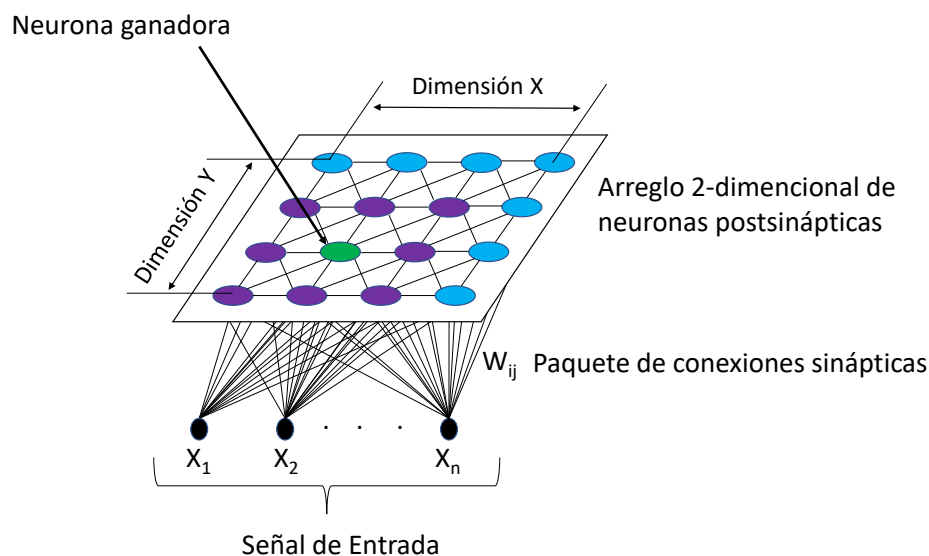
La comparación entre intervalos se basa principalmente en el parámetro de validación de *la Raíz del Error Cuadrático Medio de Validación Cruzada* (RMSECV, por sus siglas en inglés), pero también se usan otros parámetros como el coeficiente de correlación al cuadrado, la pendiente y la desviación estándar, para garantizar una descripción completa del modelo. Cabe resaltar que este método fue creado para análisis de espectros de absorción en un rango espectral de infrarrojo cercano y visible (NIR VIS). Para más información véase [105].

### 2.3. Procesamiento de datos

A continuación, se describe la teoría referente a los métodos computacionales utilizados en esta investigación para minimizar el error en la cuantificación de HbA<sub>1c</sub> y glucosa, descritas en capítulos subsecuentes.

### 2.3.1. Mapas Auto-organizados

*Self Organizing Maps* (SOM, por sus siglas en inglés) es una arquitectura ANN entrenada con un algoritmo de aprendizaje no supervisado, la cual fue creada por Teuvo Kohonen [120]. SOM tiene como objetivo encontrar patrones o características significativas en los datos de entrada y establecer una correspondencia entre estos datos y un espacio bidimensional. Con esto, se crean mapas topológicos de tal forma que datos similares activan neuronas en áreas cercanas. En la figura 2.10, se observa un conjunto de datos de entrada  $X$  con los cuales, después de crear “representaciones internas” una neurona ganadora se activará, la cual es conocida como la *unidad de mejor coincidencia* [120].



**Figura 2.10:** Modelo de mapa topológico de Kohonen con la unidad de mejor coincidencia en verde, las capas dentro del radio de vecindad en morado y los nodos en el exterior en azul [7, 11].

Este tipo de redes está diseñado para clasificar conjuntos de datos para los que ninguna organización se conoce *a priori*. SOM divide el conjunto de datos en diferentes subconjuntos (*clusters*, en inglés), con características comunes representadas por un prototipo. Con esto es posible descubrir las regularidades presentes en los datos, extraer características o agrupar patrones según su similitud [121]. Además, SOM ha tenido especial éxito en tareas que implican

señales muy ruidosas, por lo que es de especial interés en esta investigación.

### 2.3.1.1. Entrenamiento de SOM

Dado un conjunto de entrenamiento  $x$ , se quiere encontrar un conjunto de  $M$  prototipos, cada uno representado por  $w_j$  |  $j = 1, 2, 3 \dots M$  y forman un mapa de características  $\phi$ ; los valores  $w_j$  también se conocen como pesos de la red SOM.

El primer paso en el algoritmo de entrenamiento de SOM es realizar una selección aleatoria de un ejemplo de entrenamiento,  $x(0)$ , y calcular la similitud de éste con cada uno de los pesos  $w_i$  ( donde  $i$  va de 1 a el número de neuronas en SOM). Una manera de calcular dicha similitud es mediante la distancia euclidiana [122]:

$$d^2(w_i, x) = \sum_{k=1}^N (w_{ik} - x_k)^2, \text{ para } i = 1 \dots M \quad (2.24)$$

Se define como neurona ganadora  $i^*$  a aquella que obtuvo la mínima distancia a todos los  $w_i$ . Enseguida se modifican los pesos de la neurona ganadora  $i^*$  y los de sus vecinos mediante:

$$w_j(t+1) = w_j(t) + \alpha(t)(x(t) - w_j(t)), \text{ para } j \in \Lambda_{i^*}(t) \quad (2.25)$$

donde  $\Lambda_{i^*}(t)$  corresponde a una función de vecindad centrada en la neurona ganadora  $i^*$  y  $\alpha(t)$  es una función de aprendizaje [123].

Una vez entrenada, SOM puede recibir cualquier ejemplo arbitrario  $x$  y determinar la similitud de éste con todos los pesos del mapa  $\phi$ . El prototipo que representa a  $x$  será el  $w_{i^*}$  que corresponde a la *unidad de mejor coincidencia*, es decir, aquella que obtuvo la mínima distancia entre sus pesos y el ejemplo de entrada.

### 2.3.2. Extracción de características

La extracción de características es parte del proceso de reducción de dimensionalidad de los datos de entrada. Este proceso consiste en transformar datos en valores numéricos con menor dimensión, de forma tal que se conserve la información contenida en el conjunto original. El uso de procesos

de extracción de características permite obtener mejores resultados en el aprendizaje automático que los obtenidos al aplicar directamente datos sin procesar [124, 125, 126].

### 2.3.2.1. Análisis de Componentes Principales

El análisis de componentes principales (*Principal Component Analysis (PCA), en inglés*), es uno de los muchos métodos disponibles para reducir la dimensionalidad de los datos. Se puede interpretar como un método de proyección en el que los datos con  $m$  columnas (características) se proyectan en un subespacio con  $m$  columnas o menos, mientras se conserva la esencia de los datos originales [116].

El método PCA puede describirse e implementarse utilizando herramientas de álgebra lineal. Consideremos un conjunto de datos representados por una matriz  $A$  de  $n \times m$  que da como resultado una proyección de  $A$ , que llamaremos  $B$  [127]. El primer paso es calcular el promedio de cada columna de  $A$ , el cual genera una matriz que llamaremos  $F$ ; posteriormente se centran los valores en cada columna restando el valor medio de la columna, es decir,  $C = A - F$ . El siguiente paso es calcular la matriz de covarianza de la matriz  $A$  mediante la fórmula:

$$\text{cov}(X, Y) = \frac{\sum_{i=1}^n (X_i - \bar{x})(Y_i - \bar{y})}{n - 1} \quad (2.26)$$

donde  $\bar{x}$  indica el promedio de las observaciones en  $X$ ;  $Y_i$  es el  $i$ -ésimo valor de  $y$ ,  $\bar{y}$  indica el promedio de las observaciones en  $y$ ;  $n$  número total de observaciones (datos). El resultado es una matriz cuadrada de  $d \times d$  dimensiones [127].

Finalmente, se calcula la auto-descomposición de la matriz de covarianza,  $V$ ; esto implica el cálculo de vectores y valores propios [127]. Los valores propios de  $V$  corresponden a las raíces de la ecuación característica:

$$\det(V - \lambda I) = 0 \quad (2.27)$$

donde  $I$  representa la matriz identidad, y  $\lambda$  representa el valor propio de  $V$ .

El siguiente paso es formar un vector de características. Esto se construye tomando los valores propios que se desean mantener de la lista de vectores propios y se forma una matriz con estos vectores propios colocados en las

columnas [127].

$$\text{Vector de Características} = (\text{Autovector}_1, \text{Autovector}_2, \dots, \text{Autovector}_n) \quad (2.28)$$

Este es el paso final en PCA; una vez que se han elegido los vectores propios que se desean mantener en los datos y formado un vector de características, simplemente se toma la transposición del vector y se multiplica a la izquierda del conjunto de datos original, transpuesto [127].

$$\text{Datos Final} = (\text{Vector de Características}^T)(C^T) \quad (2.29)$$

Donde Vector de Características<sup>T</sup> es la matriz con los vectores propios en las columnas transpuestos de modo que los éstos ahora están en las filas. C<sup>T</sup> contiene los datos ajustados a la media y transpuestos; Datos Final representa el conjunto de datos final con elementos de datos contenidos en columnas, con sus respectivos valores por cada dimensión representados a lo largo de filas [127].

### 2.3.3. Selección de características

La alta dimensionalidad de los datos plantea desafíos. En presencia de muchas características irrelevantes, los modelos de aprendizaje tienden a sobre ajustarse y se vuelven menos comprensibles. La selección de características es otro medio eficaz para identificar características relevantes para la reducción de dimensionalidad.

Los modelos de selección de características pueden ser supervisados, no supervisados o semi-supervisados. La selección de características supervisada determina la relevancia de la característica evaluando la correlación de la característica con la clase, y sus etiquetas. La selección de características no supervisada aprovecha la variación y separabilidad de los datos para evaluar la relevancia de la característica; los algoritmos de selección de características semi-supervisados pueden usar datos etiquetados y no etiquetados. Por otra parte, los algoritmos de selección de características se dividen en tres categorías: **filtros**, **envoltorios (Wrapper)** y **modelos integrados** [128, 129].

Los algoritmos tipo **filtro** se basan en las características generales de los datos y las evalúan sin involucrar ningún algoritmo de aprendizaje. Por otra

parte, el modelo *Wrapper* requiere un algoritmo de aprendizaje predeterminado y usa su desempeño como criterio de evaluación para seleccionar características. Los algoritmos con **modelo integrado**, incorporan la selección de variables como parte del proceso, y la relevancia de las características se obtiene analíticamente a partir del objetivo del modelo de aprendizaje. Los algoritmos de selección de características con filtro y modelos integrados pueden devolver un subconjunto de características seleccionadas o las ponderaciones (medición de la relevancia de las características) de todas las características [128, 129]. A continuación abordaremos tres métodos de selección de características, dos de ellos del modelo filtro y uno del modelo *Wrapper* y todos ellos mediante algoritmos supervisados.

### 2.3.3.1. Selección de Características basada en Correlación

La selección de características basada en Correlación (*Correlation based Feature Selection (CFS)*, en inglés) asume que las características son condicionalmente independientes dada la clase y calcula una medida heurística de “mérito” de un subconjunto de características a partir de sus correlaciones, calculadas por pares [130].

La función de evaluación del subconjunto de características es la ecuación de correlación de Pearson:

$$M_s = \frac{k\bar{r}_{cf}}{\sqrt{k + k(k-1)\bar{r}_{rr}}} \quad (2.30)$$

donde  $M_s$  es la heurística “mérito” de un subconjunto de características  $S$  que contiene  $k$  características;  $\bar{r}_{cf}$  es la correlación media de clase-característica ( $f \in S$ ), y  $\bar{r}_{rr}$  es la inter-correlación media característica-característica. El numerador de la ecuación puede considerarse como una indicación de la capacidad de predicción de un conjunto de características, con respecto a una clase; el denominador indica cuánta redundancia hay entre las características [130, 12].

El cálculo  $M_s$  se realiza mediante algoritmos de búsqueda, por ejemplo “*BestFirst*” y “*GreedyStepwise*”, los cuales se describen brevemente a continuación:

- *GreedyStepwise*: Es una estrategia de búsqueda en la que se consideran cambios locales en el subconjunto de características actual. A menudo,

un cambio local es simplemente la adición o eliminación de una sola característica del subconjunto. Cuando el algoritmo considera solo adiciones al subconjunto de características, se conoce como selección hacia adelante [131, 12].

- *BestFirst*: Es una estrategia de búsqueda en la cual se evalúa cada característica hasta terminar con todas las características o se alcance la meta. Si el primer elemento es la meta entonces la búsqueda acaba, si no lo es, se elimina la primera característica y se añaden las características sucesivas, hasta terminar con todas las características. Si el camino que se explora comienza a parecer menos prometedor, esta estrategia puede dar marcha atrás a un subconjunto de características anterior más prometedor y continuar la búsqueda desde allí [131, 12].

### 2.3.3.2. RReliefF

ReliefF es un algoritmo propuesto por Kira y Rendell en 1992 [132]. La idea es estimar la calidad de los atributos de acuerdo con qué tan bien sus valores distinguen entre las instancias cercanas entre sí. En el caso de la regresión RReliefF, se penaliza a los predictores que dan valores diferentes a vecinos con los mismos valores de respuesta y se recompensa a los predictores que dan valores diferentes a vecinos con valores de respuesta diferentes. Sin embargo, RReliefF utiliza ponderaciones intermedias para calcular las ponderaciones finales del predictor [133]. Dados dos vecinos más cercanos, y considerando que:

- $W_{dy}$  es el peso de tener diferentes valores para la respuesta  $y$ .
- $W_{dj}$  es el peso de tener diferentes valores para el predictor  $F_j$ .
- $W_{dy \wedge dj}$  es el peso de tener diferentes valores de respuesta y diferentes valores para el predictor  $F_j$ .

RReliefF primero inicializa los pesos  $W_{dy}$ ,  $W_{dj}$ ,  $W_{dy \wedge dj}$  y  $W_j$  igual a cero. Luego, el algoritmo selecciona iterativamente una observación aleatoria  $x_r$ , encuentra las  $k$  observaciones más cercanas a  $x_r$  y actualiza, para cada vecino más cercano  $x_q$ , todos los pesos intermedios de la siguiente manera:

$$W_{dy}^i = W_{dy}^{i-1} + \Delta_y(x_r, x_q) \cdot d_{rq} \quad (2.31)$$

$$W_{dj}^i = W_{dj}^{i-1} + \Delta_j(x_r, x_q) \cdot d_{rq} \quad (2.32)$$

$$W_{dy \wedge dj}^i = W_{dy \wedge dj}^{i-1} + \Delta_y(x_r, x_q) + \Delta_j(x_r, x_q) \cdot d_{rq} \quad (2.33)$$

Los superíndices  $i$  e  $i - 1$  denotan el número de paso de la iteración;  $m$  es el número de iteraciones especificado por “actualizaciones”;  $\Delta_y(x_r, x_q)$  es la diferencia en el valor de la respuesta continua y entre las observaciones  $x_r$  y  $x_q$ . Sea  $y_r$  el valor de la respuesta para la observación  $x_r$ , y sea  $y_q$  el valor de la respuesta para la observación  $x_q$ , entonces:

$$\Delta_y(x_r, x_q) = \frac{|y_r - y_q|}{\max(y) - \min(y)} \quad (2.34)$$

$$\Delta_j(x_r, x_q) = \frac{|x_{rj} - x_{qj}|}{\max(F_j) - \min(F_j)} \quad (2.35)$$

RReliefF calcula los pesos predictores  $W_j$  después de actualizar completamente todos los pesos intermedios [133].

$$W_j = \frac{W_{dy \wedge dj}}{W_{dj}} - \frac{W_{dj} - W_{dy \wedge dj}}{m - W_{dy}} \quad (2.36)$$

Los pesos predictores se presentan ordenados por importancia de atributos, lo que facilita la selección de atributos. Cabe mencionar que uno de los parámetros a evaluar es el número de vecinos que el algoritmo evaluará; se recomienda iniciar con  $k = 10$ .

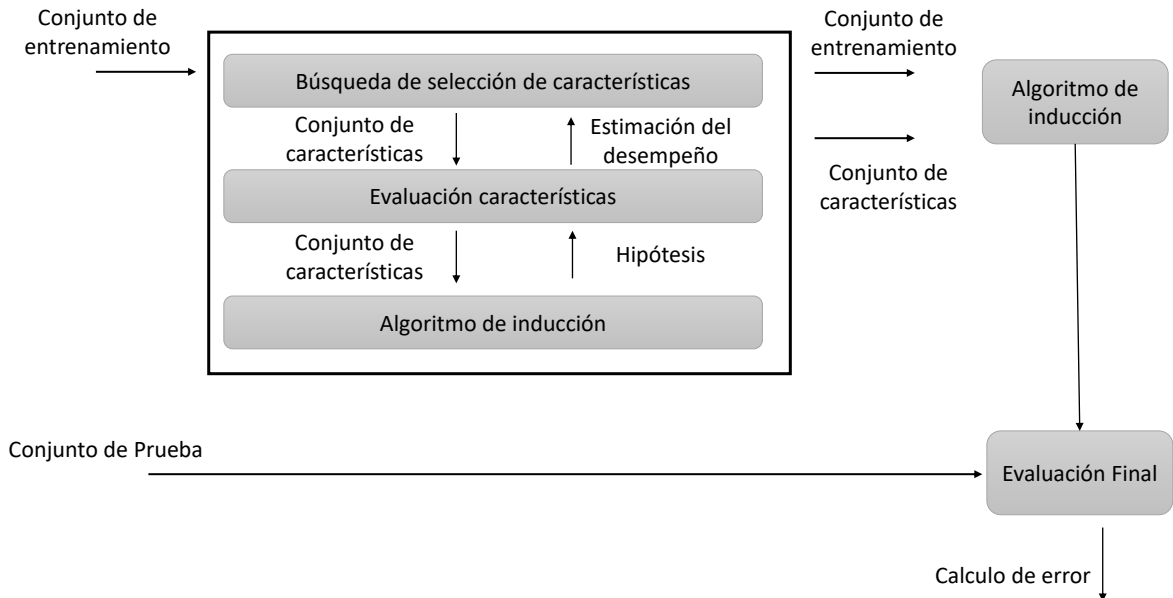
### 2.3.3.3. Método de envoltura

Como anteriormente se mencionó, existen los métodos de envoltura conocidos en inglés como “*Wrapper*”; por conveniencia se nombran de esta última forma. Este método se basa en la premisa de que para lograr el mejor rendimiento posible con un algoritmo de aprendizaje particular en un conjunto de entrenamiento particular, el método de selección de subconjuntos de características debe considerar cómo interactúan el algoritmo y el conjunto de entrenamiento [134].

La idea detrás del enfoque de envoltura, se muestra en la figura 2.11. El algoritmo de inducción, el cual se considera una caja negra, se ejecuta en



el conjunto de datos, generalmente dividido en conjuntos de entrenamiento interno y de reserva, con diferentes conjuntos de características eliminadas de los datos. El subconjunto de características con la evaluación más alta se elige como el conjunto final en el que se ejecutará el algoritmo de inducción [134].



**Figura 2.11:** Enfoque de envoltura para la selección de subconjuntos de características. El algoritmo de inducción se utiliza como una “caja negra” por el algoritmo de selección de subconjuntos [12].

Para la selección de características los métodos *Wrapper* se basan en búsquedas inteligentes como lo son: “Best First” y “GreedyStepwise” (los cuales se mencionaron en la sección de descripción de CFS); el algoritmo de inducción dependerá del problema a evaluar ya sea clasificación o regresión [134].

## 2.4. Resumen del capítulo

En este capítulo se abordaron conceptos básicos asociados a la espectroscopia Raman, que se usa en esta investigación como una técnica de sensado no invasiva. Asimismo, se cubrieron los conceptos asociados a los métodos computacionales, tales como, redes neuronales artificiales, red SOM, CFS, Wrapper, PCA, RReliefF para la cuantificación de las sustancias de interés. Cabe mencionar que en este capítulo solo se abordaron los temas de manera general; en los próximos capítulos se presentará la implementación de éstos en la solución propuesta para el problema de estudio.

---

## DESARROLLO EXPERIMENTAL

---

En este capítulo se describen los arreglos experimentales utilizados para realizar las mediciones no invasivas usando espectroscopia Raman de sustancias puras y sus disoluciones, así como mediciones en suero y mediciones *in vivo* y las implicaciones que esto conlleva. Cabe señalar que para todas las mediciones se utilizó un espectrómetro Raman Oceans Optics Raman QE65000, con resolución de 0.14-7.7nm FWHM ( $6\text{ cm}^{-1}$ ). El arreglo experimental está compuesto por una fuente láser de excitación que emite a una longitud de onda de 785 nm con potencia determinada en cada experimento y una punta de prueba InPhotonics Raman de RIP-RPS-785<sup>TM</sup>.

### 3.1. Mediciones de Glucosa

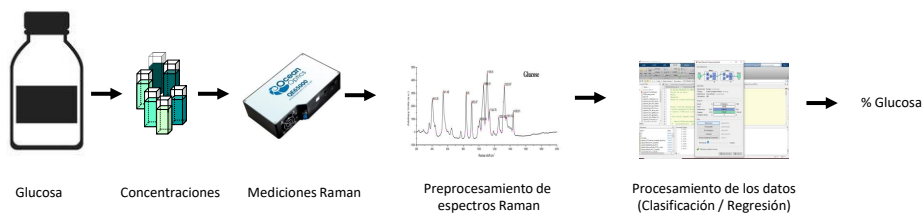
Se realizaron mediciones de glucosa con la finalidad de caracterizar dicha sustancia usando espectroscopia Raman y localizar experimentalmente sus picos representativos, En este caso se utilizó una potencia de 100 mW. Para estabilizar el sistema, la adquisición de los espectros comenzó 10 minutos después de que el láser fue encendido; las mediciones se tomaron cada 25 s (10 mediciones por muestra), ajustando la distancia de trabajo en cada medida, hasta obtener la mayor intensidad de señal que permitió mejorar la calidad de los espectros.

La glucosa anhidra se adquirió de Sigma-Aldrich Corporation, St Louis, Missouri, EE. UU., SKU: G8270. Se prepararon mezclas con porcentajes de de 100 %, 50 %, 40 %, 30 %, 20 % y 10 %, para su preparación se determinó el porcentaje total de soluto en la disolución considerando la densidad del soluto ( $\rho_{\text{solute}}$ ) y masa del soluto ( $m_{\text{solute}}$ ), ambos parámetros constantes; el volumen

de disolución ( $V_{\text{disolucion}}$ ) se calcula como [135]:

$$V_{\text{disolucion}} = \frac{100 * m_{\text{solute}}}{\frac{m_{\text{solute}}}{V_{\text{solute}}}} - \frac{m_{\text{solute}}}{\rho_{\text{solute}}} \quad (3.1)$$

De las concentraciones mencionadas anteriormente se realizaron mediciones en diferentes días, lo que implica que algunas condiciones ambientales, tales como temperatura y humedad no fueron controladas. La figura 3.1 muestra la secuencia utilizada para la determinación de porcentajes de glucosa, en la cual se observa como primer paso la preparación de las concentraciones, posteriormente la adquisición de espectros Raman, el preprocesamiento de la señal, es decir, eliminación de fluorescencia y suavizado, seguido por el uso de las redes neuronales artificiales ya sea como clasificación o regresión y por último la obtención del resultado, para más información véase [136].



**Figura 3.1:** Configuración experimental y análisis de datos.

## 3.2. Mediciones de HbA<sub>1c</sub>

Con el objetivo de analizar HbA<sub>1c</sub> en pacientes diabéticos, como una primera aproximación se realizaron mediciones de HbA<sub>1c</sub> pura, en las cuales se utilizó una potencia de 10 mW, ésta se obtuvo por pérdida de acoplamiento, es decir, cambiando la separación entre dos fibras enfrentadas, esto permite modular la potencia recibida [137].

La HbA<sub>1c</sub> en polvo fue adquirida mediante Sigma Aldrich (ID: IRMMIFCC466). El polvo puro se consideró como una concentración del 100 %; además, se prepararon dos soluciones utilizando 10 µl y 20 µl de agua destilada en 1 mg de HbA<sub>1c</sub>, correspondientes a concentraciones de 9,09 % y 4,76 % respectivamente; estas muestras son similares a las concentraciones

encontradas en el cuerpo humano [138]; se realizaron diez mediciones para cada concentración. Las soluciones se prepararon utilizando la ecuación 3.2 [139]; la masa de soluto ( $m_{\text{solute}}$ ) se estableció en 1 mg y la densidad del agua ( $\rho_{\text{agua}}$ ) se utilizó como 1 g/cm<sup>3</sup> [140]; el volumen de agua ( $v_{\text{agua}}$ ) se varió de [10 µl: 20 µl].

$$\text{Concentración} = \frac{m_{\text{solute}}}{m_{\text{solute}} + (\rho_{\text{agua}} * v_{\text{agua}})} * 100 \quad (3.2)$$

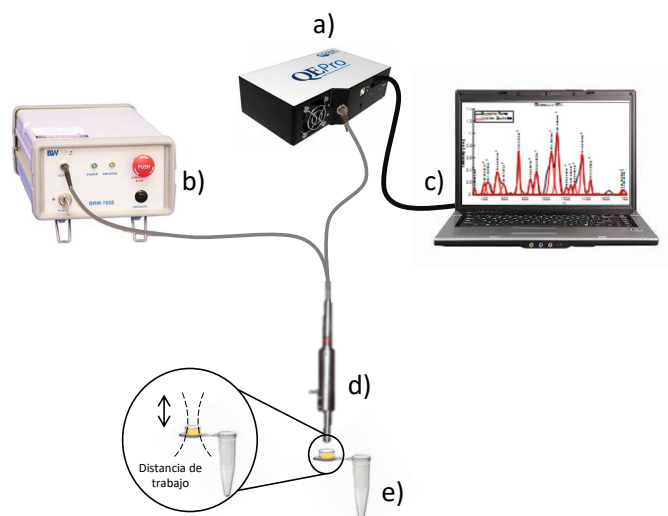
Para estas mediciones se utilizó un contenedor cilíndrico de polipropileno color negro con la finalidad de obtener mejores resultados en la adquisición de los espectros Raman, puesto que si se utilizaba un contenedor transparente el detector del espectrómetro presentaba saturación y con ello no se obtenían medidas adecuadas. Para más información de las mediciones realizadas véase [1].

### 3.3. Mediciones de suero humano

La finalidad de esta investigación es utilizar esta metodología en la detección de HbA1c y glucosa en hospitales, para ello se realizaron mediciones experimentales en suero humano, como primer paso para determinar los parámetros de medición consistió en realizar una prueba preliminar con 8 voluntarios (integrantes del Grupo de Instrumentación y Óptica Biomédica, GIOB). Una vez determinados estos parámetros se proporcionaron 50 sueros de pacientes diabéticos por el Instituto Mexicano del Seguro Social (IMSS). Las medidas de referencia se tomaron utilizando el método de glucosa oxidasa y los valores de glucosa oscilaron entre 80 mg/dL y 439 mg/dL.

Se utilizó una potencia de 100 mW, las mediciones se tomaron durante 60 segundos y se realizaron diez repeticiones por suero. Así mismo, se estableció la distancia entre la punta Raman y la muestra a 7 mm como sugiere el manual, también se experimentó con diferentes distancias hasta encontrar la máxima amplitud en la señal Raman, siendo la distancia mencionada la que presentó la máxima amplitud.

Como se observa en la figura 3.2 el contenedor utilizado fue un recipiente para micro centrifuga de polipropileno, sin embargo, por experimentación



**Figura 3.2:** Configuración experimental: a) Espectrómetro Raman QE65000, b) Fuente láser de 785 nm, c) Computadora, d) Punta de prueba Raman, e) Recipiente de muestra.

posterior, se observó que este tipo de contenedores generan señales que no son propias de la muestra.

Considerando que las mediciones anteriores se realizaron en un contenedor que generó señales no propias del suero humano, se realizó una segunda medición de sueros, proporcionado por el laboratorio de análisis clínicos “Diagnoscentro”, ubicado en Tuxtla Gutiérrez, Chiapas, se analizaron 46 sueros de voluntarios (no diabéticos, pre-diabéticos y diabéticos), aproximadamente 30  $\mu\text{l}$  por suero fueron proporcionados, para estas mediciones se utilizó un contenedor de aluminio 6067 diseñado y fabricado en el Grupo de Instrumentación y Óptica Biomédica (GIOB), con la finalidad de evitar contribuciones ajenas a la muestra, con ello también se disminuyó el tiempo de integración a 30 segundos, los valores de glucosa oscilaron entre 56 mg/dL - 400 mg/dL, de igual manera estos valores se obtuvieron mediante el método de glucosa oxidasa [141].

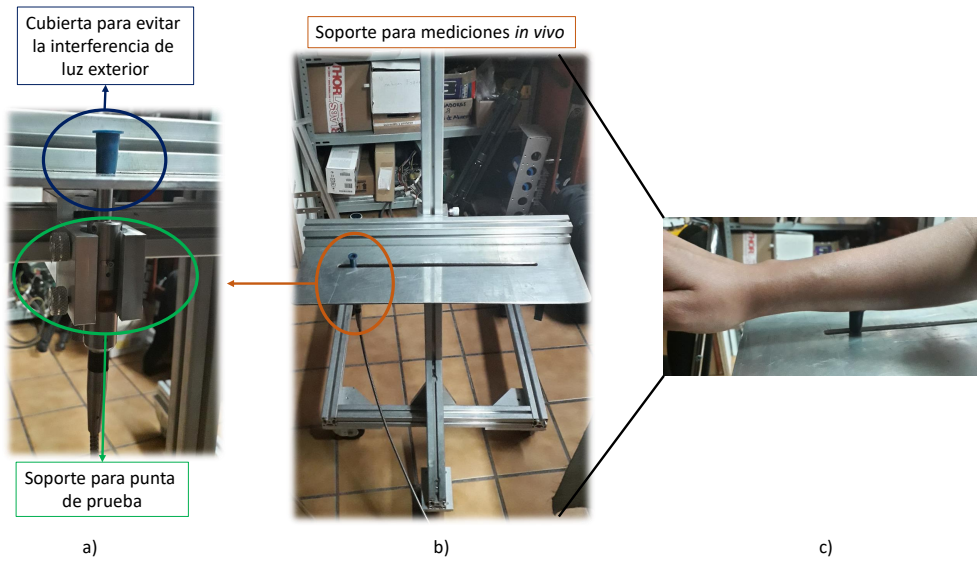
### 3.4. Mediciones *in vivo*

La metodología desarrollada no solamente permite analizar muestras *in vitro*, también es posible analizar con ella muestras *in vivo* para lo cual como primer paso se consideró el espacio adecuado para la adquisición de espectros Raman de los voluntarios, puesto que estas mediciones se realizarían fuera del laboratorio. Por lo anterior se utilizó un soporte prediseñado y fabricado en el GIOB, el cual es desarmable para su fácil transportación. Sin embargo, este soporte carecía de ciertas condiciones para realizar las mediciones, así que se diseñó en Solid Works<sup>®</sup> algunos aditamentos para sostener la punta de prueba, así como un soporte para el apoyo del brazo del voluntario; también se implementó un sistema para asegurar la no interferencia de luz externa en la zona de prueba, esto se observa en la figura 3.3; en a) se muestra el soporte y la punta de prueba Raman, así como la cubierta para evitar que la luz exterior sea captada por el detector; b) muestra el soporte completo, y c) un ejemplo del posicionamiento del antebrazo en el soporte. Un punto importante en esta implementación fue el sistema para la no interferencia de luz externa, en el cual se probaron 3 distintos modelos, de los cuales dos al colocarlos sobre la punta de prueba y adquirir espectros Raman de alcohol (se eligió esta sustancia porque sus picos Raman están reportados en la literatura [142]) presentaron picos Raman ajenos a la muestra analizada, es decir, los picos que presentó la señal eran parte de este aditamento, por lo cual se descartaron para su uso, el aditamento elegido al tener una forma cónica, la luz no tuvo interacción con las paredes del mismo, siendo adecuado para su uso en esta aplicación.

Es importante señalar que para las mediciones realizadas se contó con la firma de consentimiento informado de cada voluntario, véase apéndice C.1 para observar el formato.

#### 3.4.1. Selección de parámetros

Un parámetro importante a considerar para la realización de mediciones *in vivo* es la potencia del láser, la cual se calculó considerando la norma ANSI Z136.1-2007 [143], esta menciona el uso de láseres en espectroscopia, proporcionando una clasificación de láseres en función de sus efectos biológicos



**Figura 3.3:** Prototipo para mediciones *in vivo*.

y la máxima exposición permitida (MPE, por sus siglas en inglés) para uso en ojos y piel, en dicha norma se debe considerar la distancia focal, el tamaño del spot del láser, la apertura numérica (véase tabla III.1 de la norma), la longitud de onda del láser, y un aproximado del tiempo de exposición, el cálculo de MPE se presenta a continuación:

$$MPE = 0,2C_A \quad (3.3)$$

En donde  $C_A$  depende de la longitud de onda ( $\lambda$ ) de excitación y está dada por:

$$C_A = 10^{2(\lambda-0,07)} \quad (3.4)$$

$$C_A = 10^{2(0,785-0,07)} = 1,48 \text{ para } \lambda = 788 \text{ nm} \quad (3.5)$$

$$MPE = (0,2)(1,48) = 0,30 \frac{W}{cm^2} \quad (3.6)$$

Es importante considerar el tamaño de spot como se mencionó anteriormente para eso realizamos el cálculo de área irradiada considerando el tamaño del spot  $\phi$  mencionado en el manual de la punta de prueba, en este caso es  $\phi = 189,6\mu m$ .

$$A = \pi * r^2 \quad (3.7)$$



$$A = \pi * (0,00948\text{cm}^2) = 0,0002823361\text{cm}^2 \quad (3.8)$$

Considerando el cálculo de área de la ecuación 3.8, es posible conocer la exposición máxima permitida en esa área mediante:

$$\text{MPE}(A) = 0,30 \frac{\text{W}}{\text{cm}^2} * 0,0002823361\text{cm}^2 = 0,0835207\text{mW} \quad (3.9)$$

En 3.9 la potencia calculada según la norma es para un tiempo de 500 minutos ( $3 \times 10^4$  segundos), a continuación mediante el uso de una proporcionalidad inversa [Falta Ecuación] se calcula MPE para diferentes tiempos de exposición, los resultados se resumen en la tabla 3.1.

**Tabla 3.1:** Resultado del cálculo de máxima exposición permitida (MPE) para diferentes tiempos.

| Tiempo      | MPE(A) en mW |
|-------------|--------------|
| 60 segundos | 41.46        |
| 50 segundos | 50.11        |
| 40 segundos | 62.64        |
| 30 segundos | 83.52        |
| 20 segundos | 100.22       |
| 15 segundos | 167.05       |

Aunado a esto se realizaron pruebas preliminares usando espectroscopia Raman en el dedo índice de 3 voluntarios, con la finalidad de evaluar estos tiempos en función de parámetros como desviación estándar y relación señal a ruido (SNR, por sus siglas en inglés); en donde para la elección de la mejor potencia se consideró la menor desviación estándar y la mayor SNR de los experimentos. Con lo cual se determinó una potencia de **60 mW** con un tiempo de integración de **30 segundos**.

### 3.4.2. Selección de regiones corporales

Bajo la hipótesis de que la sangre circula en todo el cuerpo y con ella la glucosa y hemoglobina, determinamos la detección de glucosa en regiones corporales provistas de una gran cantidad de vasos sanguíneos. La determinación de las regiones corporales a utilizar se realizó de manera experimental mediante la mediciones de 21 voluntarios no diabéticos, en un rango de edad de los 24 a los 58 años (media 33 años), las mediciones se realizaron en la frente, lóbulo de la oreja derecha, tenar mano derecha, muñeca mano derecha, dedo índice mano derecha y antebrazo derecho, de estos voluntarios se obtuvieron los espectros Raman, así como el valor de glucosa utilizando un glucómetro One Touch Ultra 2, el cual funciona mediante tiras reactivas que contienen glucosa oxidasa, una enzima que reacciona con la glucosa en una gota de sangre, y una interfaz con un electrodo dentro de un medidor [144]. Los parámetros que se consideraron para determinar la zona de medición fueron: la relación señal a ruido (SNR) de cada región, en la cual es preferible valores altos y/o mayor a 3 con la finalidad de poder cuantificar la sustancia bajo análisis [92], así como los resultados obtenidos de la aproximación de los niveles de glucosa obtenidos por el glucómetro usando redes neuronales artificiales y la diferencia existente entre los picos Raman encontrados en las sustancias puras y los encontrados en los espectros *in vivo*. Los resultados del cálculo de la SNR se muestran en la tabla 3.2, dicho valor se calculó para las dos sustancias analizadas, glucosa y HbA<sub>1c</sub>.

**Tabla 3.2:** Relación señal a ruido (SNR) de los 21 voluntarios.

| Región corporal | SNR glucosa<br>(pico 1106 cm <sup>-1</sup> ) | SNR HbA <sub>1c</sub><br>(pico 1230 cm <sup>-1</sup> ) |
|-----------------|--|--|
| Antebrazo       | <b>5.88</b>                                  | 5.56   |
| Dedo            | 3.92   | 4.94   |
| Frente          | 3.27   | 4.85   |
| Lóbulo          | 3.41   | 4.77   |
| Muñeca          | 4.07   | <b>5.94</b>  |
| Tenar           | 2.82   | 4.49   |

En el caso de la diferencia de picos se realizó una comparación de los picos

obtenidos en la sustancia pura de glucosa con los obtenidos en las mediciones *in vivo*, la tabla 3.3 presenta los resultados obtenidos en donde en promedio el tenar obtuvo una diferencia de  $3.53 \pm 0.86 \text{ cm}^{-1}$ , el antebrazo de  $3.58 \pm 1.57 \text{ cm}^{-1}$ , la muñeca de  $3.8 \pm 1.29 \text{ cm}^{-1}$ , el dedo de  $3.88 \pm 1.20 \text{ cm}^{-1}$ , el lóbulo de la oreja de  $4.76 \pm 1.75 \text{ cm}^{-1}$ , la frente de  $5.64 \pm 2.40 \text{ cm}^{-1}$ .

**Tabla 3.3:** Análisis de diferencia de picos representativos de glucosa.

| Regiones corporales | Glucosa                          |                       |                       |                      |                      |
|---------------------|----------------------------------|-----------------------|-----------------------|----------------------|----------------------|
|                     | Picos Raman ( $\text{cm}^{-1}$ ) |                       |                       |                      |                      |
|                     | 1105 $\text{cm}^{-1}$            | 1325 $\text{cm}^{-1}$ | 1059 $\text{cm}^{-1}$ | 835 $\text{cm}^{-1}$ | 906 $\text{cm}^{-1}$ |
| Antebrazo           | 4.32                             | 4.55                  | 4.73                  | 0.91                 | 3.41                 |
| Dedo índice         | 4.59                             | 4.41                  | 4.36                  | 1.73                 | 4.32                 |
| Frente              | 7.73                             | 5.00                  | 6.91                  | 1.73                 | 6.86                 |
| Lóbulo oreja        | 6.73                             | 5.27                  | 4.36                  | 2.00                 | 5.46                 |
| Muñeca              | <b>4.77</b>                      | <b>4.77</b>           | <b>4.46</b>           | <b>1.77</b>          | <b>3.23</b>          |
| Tenar               | 4.05                             | 3.82                  | 4.00                  | 2.00                 | 3.82                 |

En la tabla 3.4 se muestra la diferencia entre picos de la sustancia pura de HbA1c con lo encontrado en los espectros *in vivo*, en donde se observa que en promedio la muñeca tiene una diferencia de  $3.32 \pm 1.02 \text{ cm}^{-1}$ , el dedo índice  $3.69 \pm 1.01$ , el tenar de  $4.03 \pm 0.98$ , el antebrazo de  $4.12 \pm 1.19$ , el lóbulo de la oreja de  $4.70 \pm 2.08$ , la frente de  $4.94 \pm 2.57$ .

**Tabla 3.4:** Análisis de diferencia de picos representativos de HbA1c.

| Regiones corporales | HbA1c                            |                       |                      |                       |                       |
|---------------------|----------------------------------|-----------------------|----------------------|-----------------------|-----------------------|
|                     | Picos Raman ( $\text{cm}^{-1}$ ) |                       |                      |                       |                       |
|                     | 1230 $\text{cm}^{-1}$            | 1220 $\text{cm}^{-1}$ | 985 $\text{cm}^{-1}$ | 1313 $\text{cm}^{-1}$ | 1319 $\text{cm}^{-1}$ |
| Antebrazo           | 4.64                             | 3.14                  | 3.00                 | 3.96                  | 5.91                  |
| Dedo índice         | 3.86                             | 3.77                  | 2.00                 | 4.09                  | 4.73                  |
| Frente              | 4.27                             | 2.86                  | 2.82                 | 5.77                  | 9.00                  |
| Lóbulo oreja        | 5.14                             | 1.91                  | 3.46                 | 5.68                  | 7.32                  |
| Muñeca              | <b>4.36</b>                      | <b>2.73</b>           | <b>2.00</b>          | <b>3.23</b>           | <b>4.32</b>           |
| Tenar               | 4.18                             | 4.46                  | 2.36                 | 4.23                  | 4.96                  |

Así mismo para la elección de la zona corporal, se realizó una regresión preliminar usando redes neuronales artificiales, en donde se probaron diferentes rangos espectrales, siendo estos:  $R_1 = 800\text{-}1000 \text{ cm}^{-1}$ ,  $R_2 = 1000\text{-}1200$

$\text{cm}^{-1}$ ,  $R_3 = 1200\text{-}1400 \text{ cm}^{-1}$ . Los valores de glucosa variaron de 86 a 136 mg/dL todos los sujetos fueron no diabéticos, los resultados se presentan en la tabla 3.5.

**Tabla 3.5:** Análisis de diferencia de picos representativos de glucosa.

| Región corporal | Error absoluto valor real - predicho (mg/dL) |       |       |          |
|-----------------|--|-------|-------|----------|
|                 | R1   | R2    | R2    | promedio |
| Frente          | 11.59  | 7.95  | 10.86 | 10.13    |
| Muñeca          | 7.29   | 10.28 | 6.75  | 8.10     |
| Tenar           | 10.93  | 12.46 | 11.00 | 11.46    |
| Lóbulo          | 10.08  | 9.97  | 8.19  | 9.42     |
| Dedo            | 8.41   | 8.84  | 9.02  | 8.75     |
| Antebrazo       | 8.50   | 12.22 | 8.56  | 9.76     |

De acuerdo con lo anterior se eligieron las zonas del **antebrazo, muñeca y dedo**, cabe resaltar que el antebrazo ya ha sido usado en otras investigaciones [70, 66, 103], de igual manera el uso de la región del dedo ya ha sido reportada antes [99], sin embargo del uso de la zona de muñeca no se encontraron reportes anteriores.

### 3.4.3. Adquisición de espectros Raman *in vivo*

La adquisición de los espectros Raman *in vivo* se llevó acabo en el laboratorio de análisis clínicos “Diagnoscentro” ubicado en Tuxtla Gutiérrez, Chiapas, en donde fue proporcionado un lugar exclusivo para colocar toda la instrumentación y realizar las mediciones; esta colaboración se realizó mediante el Centro para la Prevención y Control del Cáncer (CEPREC), el laboratorio Diagnoscentro e INAOE.

Se analizaron 46 voluntarios, de los cuales 4 fueron no diabéticos, 10 pre-diabéticos y 32 diabéticos; 16 hombres y 30 mujeres en un rango de 27 a 87 años (media 51 años), peso entre 45 - 125 kg (media 74 kg) y estatura entre 142 - 180 cm (media 160 cm). Las condiciones ambientales fueron: humedad  $63.09 \pm 5.25 \%$  y temperatura  $19.76 \pm 1.02 \text{ }^{\circ}\text{C}$ .

Como se estableció anteriormente las zonas determinadas para las mediciones fueron el antebrazo, la muñeca y el dedo índice, con potencia de 60 mW, tiempo de integración de 30 segundos, 2 boxcar de promediación. Del

antebrazo y muñeca se tomaron 10 mediciones por voluntario, mientras que del dedo solo 3 mediciones. Así mismo se realizaron mediciones antropométricas tales como, ancho de muñeca, ancho de antebrazo y largo del antebrazo, así como temperaturas iniciales y finales de cada medición en las diferentes parte del cuerpo, en la figura 3.4 se muestra el proceso para la medición *in vivo*, en el cual se observa como primer paso la calibración del equipo, así como la medición de la potencia inicial (60 mW), posteriormente se mide la temperatura ambiental, y la temperatura corporal de la región analizada; a continuación el voluntario coloca la región corporal en la punta de prueba (azul) en esta diagrama se muestra el dedo como ejemplo y por último se realizan la adquisición de espectros Raman y se mide la temperatura final de la región corporal analizada.

Los valores de referencia fueron determinados por el laboratorio “Diagnoscentro” mediante la extracción sanguínea en cada voluntario mediante el método de “afinidad de boronato” [145] obteniendo como resultado valores entre 5.2 % y 14 % para HbA<sub>1c</sub>. En el caso de glucosa los valores van de 56 mg/dL a 400 mg/dL y el método de obtención fue mediante “glucosa oxidasa” [141].



**Figura 3.4:** Proceso de mediciones *in vivo* mediante espectroscopia Raman.

### 3.5. Resumen del capítulo

En este capítulo se presentaron los arreglos experimentales para la adquisición de espectros Raman de glucosa y HbA<sub>1c</sub> puras y disoluciones, así como para suero y espectros de mediciones *in vivo* en antebrazo, muñeca y dedo de voluntarios diabéticos y no diabéticos; el parámetro de potencia varió para cada sustancia analizada, así como el contenedor en el que se midieron. En el caso de las mediciones *in vivo* se calculó la potencia máxima y el tiempo de integración necesarios para no dañar el tejido humano conforme a normas específicas. Así mismo se diseñaron y fabricaron aditamentos necesarios para realizar dichas mediciones. En el siguiente capítulo, se presentarán los resultados de implementar diversos métodos computacionales para la cuantificación de HbA<sub>1c</sub> y glucosa en los diferentes espectros Raman adquiridos con la metodología descrita en este capítulo.

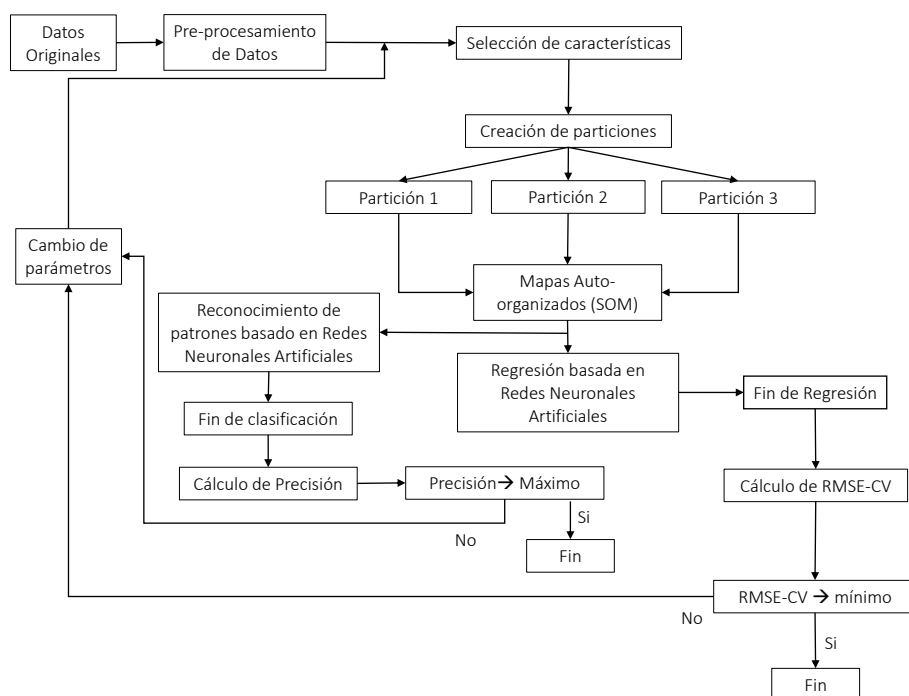
---

## PROCESOS DE MEDICIÓN

---

En este capítulo se describen los procesos utilizados para diseñar el sistema capaz de aproximar concentraciones de HbA<sub>1c</sub> y glucosa, los cuales están basados en algoritmos de selección de características, mapas auto-organizados SOM y ANN. Las aproximaciones se llevaron a cabo de dos formas: usando regresiones y usando clasificadores; ambos tipos fueron evaluado en varias bases de datos, las cuales se describen en la sección 3. El capítulo se organiza de la siguiente manera: primero se define el problema de regresión y se detallan los selectores de características utilizados; la segunda parte describe los algoritmos para diseñar las ANN con máximo desempeño. En la tercera parte se detalla el proceso de clasificación y finalmente presentamos un análisis espectral.

En la figura 4.1 se muestra el proceso de diseño de los sistemas de regresión y clasificación basados en ANN. Una vez adquiridos los espectros, se utilizan algoritmos de suavizado y eliminación de fluorescencia, descritos en el capítulo 2. El segundo paso es la selección de características, seguido de la generación de prototipos espectrales utilizando mapas auto-organizados SOM; los prototipos alimentan al modelo de regresión o clasificación. Los modelos de regresión se utilizan para aproximar los valores exactos de HbA<sub>1c</sub>, mientras que los clasificadores solo distinguen entre 3 clases: no diabético, diabético y pre-diabético. El diseño de los modelos neuronales se lleva a cabo utilizando validación cruzada de 3 particiones.



**Figura 4.1:** Diseño de sistemas de regresión y clasificación para la determinación de HbA1c y glucosa.

## 4.1. Regresión

Para realizar la regresión se utilizaron los espectros Raman medidos en tres partes diferentes del cuerpo (antebrazo, muñeca y dedo), de los cuales se obtuvieron nueve mediciones efectivas por voluntario (46 voluntarios en total). En casos de antebrazo y muñeca, se obtuvieron un total de 414 mediciones en un rango de  $200$  a  $1800\text{ cm}^{-1}$ , en el caso del dedo se tomaron 3 medidas por voluntario, un total de 138 mediciones. De las mediciones obtenidas el ruido de disparo y la fluorescencia fueron eliminados de cada espectro.

El análisis se implementó usando dos rangos espectrales, el primero de  $200$  a  $1800\text{ cm}^{-1}$ , equivalente a 788 valores utilizados como características; el segundo rango está localizado de  $600$  a  $1600\text{ cm}^{-1}$  considerando los productos finales de glicación avanzada, que ocurre en esta región [103], este rango tiene 512 características. Los espectros se filtraron con la finalidad de eliminar el ruido de disparo y la fluorescencia, para ello se utilizaron dos propuestas del Grupo de Instrumentación y Óptica Biomédica (GIOB) del INAOE, la primera



consiste en wavelets para eliminar el ruido de disparo y la otra en ajuste de polinomios de Zernike en conjunto con algoritmos genéticos para eliminar la fluorescencia. Cabe destacar que estos algoritmos se utilizaron en todos los casos analizados, es decir, en la parte del pre-procesamiento espectral, sin modificar sus parámetros. Los parámetros utilizados para cada uno de los métodos utilizados son los siguientes:

- En el caso de wavelets se utilizó la familia wavelets Daubechies, las cuales son ortogonales, se realizó una descomposición en wavelets de la señal a un nivel 5 usando symlet 6 y después se estimó la desviación estándar del ruido esto se obtiene del detalle de los coeficientes al nivel 1, correspondiente al índice de nodo 2. Se usó el **wpbmpen** para seleccionar el umbral global para quitar el ruido de la señal, usando el parámetro recomendado alpha 2. Luego se usó **wpdencmp** para remover el ruido de la señal utilizando un umbral por encima del umbral suave y manteniendo la aproximación.
- Para la eliminación de fluorescencia se utilizaron los primeros 11 polinomios de Zernike, estos varían en función del espectro analizado, para el ajuste de esos polinomios se utilizaron algoritmos genéticos implementados Matlab usando la caja de herramientas Global Optimization y un algoritmo lamarckiano en conjunto con uno híbrido de optimización. En el cual se utilizaron las funciones **ga**, **gaoptimset**, y **fmicom** en donde se utilizó un tamaño de población 200, un número de generaciones de 200, elitismo de 20, el número máximo de generaciones y generaciones de estancamiento infinito, así como tiempo de estancamiento infinito y los parámetros de la función a evaluar que en este caso se limitó la búsqueda de los coeficientes de los polinomios de Zernike de -10 a 10.

Se usó la caja de herramientas Matlab 2019b “Deep Learning”, específicamente la aplicación “Neural Net Fitting” para el diseño de la regresión y “Neural Pattern Recognition” para la clasificación [146]. El código generado por la herramienta se usó para encontrar los valores de algunos hiper-parámetros que permitieran obtener las redes con mejor desempeño: el número de nodos ocultos se varió de 5 a 20, los algoritmos de entrenamiento probados fueron el Gradiente conjugado escalado y Levenberg-Marquardt; las

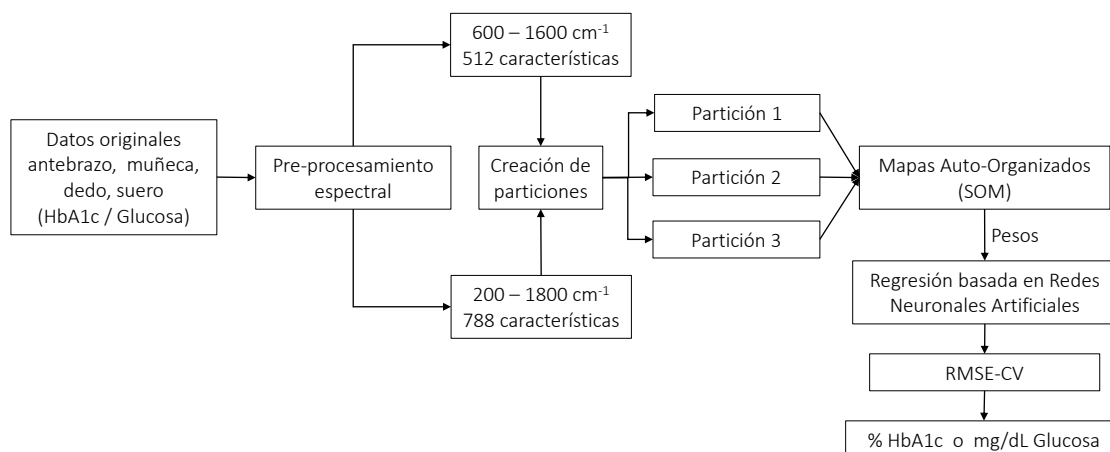
funciones de activación involucradas fueron la sigmoidea tangente hiperbólica y lineal, en la capa oculta, las cuales se probaron con la finalidad de elegir la que mejor resultados produjera, en el caso de la capa de salida se utilizó la lineal en el caso de regresión y la softmax para el caso de clasificación. Cabe hacer notar que acorde con lo encontrado en el estado del arte [147, 148], se decidió usar “Levenberg-Marquardt” como algoritmo de entrenamiento para regresión, y el de “Gradiente conjugado escalado” para clasificación, sin embargo en el caso de clasificación a pesar de que es posible ocupar otros algoritmos, que de igual manera acorde a lo reportado en la literatura, tienen un buen desempeño en problemas de reconocimiento de patrones, no se probaron debido a que la finalidad de este trabajo fue la cuantificación y no la clasificación. Los parámetros para cada uno de estos algoritmos se encuentran en el apéndice E.1. Cabe resaltar que la tasa de aprendizaje no se modificó considerando que estos algoritmos de entrenamiento no la usan directamente [149, 150, 151].

#### 4.1.1. Mapas auto-organizados (SOM)

Los mapas auto-organizados SOM son utilizados en este trabajo para generar prototipos que mejoran la representación de los espectros Raman. Como se mencionó anteriormente, SOM es una ANN capaz de generar agrupamiento, conteniendo en los pesos de cada nodo la mejor representación de los integrantes de cada grupo, o sea el prototipo correspondiente. La idea fundamental detrás del uso de SOM es obtener prototipos que representen a los diferentes patrones presentes en los espectros Raman y utilizar el prototipo, en vez del dato original, como el vector de entrada a los procesos de clasificación o regresión. El objetivo de este proceso es mejorar la señal, sin afectar la dimensionalidad de los datos.

La figura 4.2 muestra los pasos para determinar las concentraciones de glucosa y HbA<sub>1c</sub> incluyendo el mejoramiento con SOM. Se usó la caja de herramientas *Deep Learning* de Matlab 2019b para generar un código de ejecución de SOM modificable. En el diagrama se muestran los diferentes dos rangos espectrales como entrada de SOM; una vez entrenada la SOM, los pesos del prototipo correspondientes a un dato particular son la entrada para una ANN multicapa. Un hiper-parámetro importante a considerar en SOM es el tamaño de los mapas; en los experimentos reportados aquí, este valor se

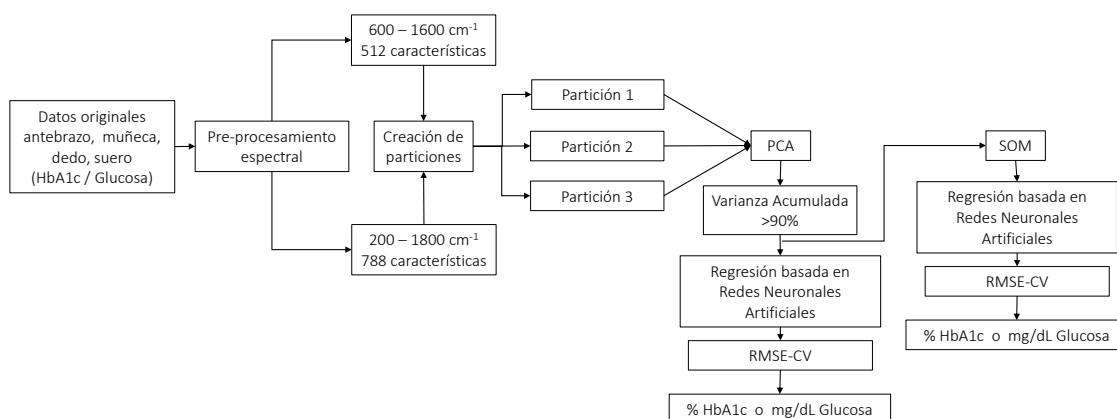
generó de forma experimental encontrándose que un mapa de  $10 \times 10$  produce resultados satisfactorios en este contexto.



**Figura 4.2:** Sistema de medición utilizando mapas auto-organizados SOM.

#### 4.1.2. Análisis de Componentes Principales

Para seleccionar el número de componentes principales (PC, por sus siglas en inglés) que se utilizan como entrada a la ANN multicapa, se consideró el porcentaje de varianza acumulada de los datos, el cual se determinó alrededor del 90 %, esto es, el 90 % de los datos totales están representado con este número de componentes principales. El análisis incluyó el cálculo de 100 PC's sobre los dos rangos espectrales involucrados en este trabajo. Usando PC se obtuvo se redujo la dimensionalidad de los datos de 788 o 512 características de entrada a 100 características. Esto implica una mejora en el tiempo de cómputo; los PC se calcularon utilizando el comando **pca** de Matlab 2019b. La Figura 4.3 muestra el proceso utilizado para calcular las concentraciones de HbA1c y glucosa incluyendo PC's; se muestran los dos rangos espectrales, la creación de las 3 particiones para la validación cruzada, el cálculo de PC's y la ANN multicapa para la cuantificación de las sustancias analizadas. Nótese que se mantiene el uso de SOM.



**Figura 4.3:** Sistema de medición utilizando Análisis de Componentes Principales y SOM.

### 4.1.3. Selección de características basada en correlación

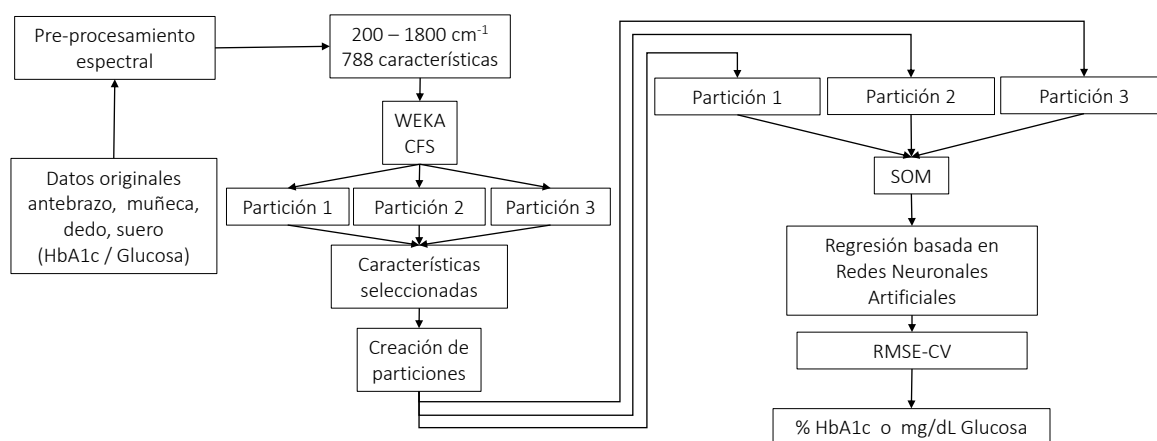
Para el diseño de esta parte del sistema de medición se usó el software libre WEKA de la Universidad de Waikato, Nueva Zelanda [152]. Este es un software en el que es posible analizar datos, por ejemplo, selección de características, clasificación, agrupación, etc. Para usar WEKA se requiere que los datos tengan la representación que se muestra en la tabla 4.1, en la cual se describe un par ordenado, donde  $m$  representa el número de característica,  $n$  el número de observaciones,  $m + 1$  corresponde a la fila que representa la salida del sistema, es decir, el valor de la concentración de glucosa o HbA1c. Los datos son manejados a través de archivos con formato “.csv”.

**Tabla 4.1:** Representación de datos de entrada para WEKA.

| 1        | 2        | 3        | ...  | m        | m+1     |
|----------|----------|----------|------|----------|---------|
| Dato 1,1 | Dato 1,2 | Dato 1,3 | .... | Dato 1,m | Valor 1 |
| Dato 2,1 | Dato 2,2 | Dato 2,3 | ...  | Dato 2,m | Valor 2 |
| .        | .        | .        | .    | .        | .       |
| .        | .        | .        | .    | .        | .       |
| .        | .        | .        | .    | .        | .       |
| Data 1,m | Data n,2 | Data n,3 | ...  | Dato n,m | Valor m |

La selección de características se aplicó solamente en los datos que utilizan el rango espectral uno, puesto que lo que se busca es determinar qué características de todo el conjunto de datos son las mejores para representar

el problema. Como se mencionó antes, CFS basa sus resultados en la búsqueda heurística y algunos algoritmos de búsqueda son: “BestFirst” y “GreedyStepwise”. De forma similar a los otros diseños, se utilizó una validación cruzada de tres dobleces. El algoritmo de selección de WEKA proporciona como resultado las veces que se utilizó una característica determinada por partición; en nuestro caso se seleccionaron las características que se usaron al menos en una de las particiones. La figura 4.4 muestra el proceso de selección de características. Las características seleccionadas por WEKA se usan como entrada para la red SOM; una vez generados los pesos, éstos alimentan la ANN multicapa para la obtención de la concentración de HbA<sub>1c</sub> y glucosa. Se espera que el número de características seleccionadas sea menor que el número de características en la entrada a la selección.

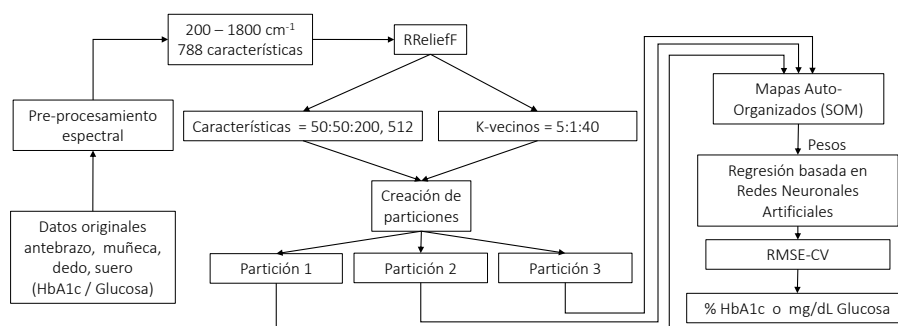


**Figura 4.4:** Sistema de medición utilizando selección de características basada en correlación (CFS).

#### 4.1.4. Selección de características usando RReliefF

RReliefF ordena todas las características de entrada de acuerdo con la importancia que tienen con respecto a la variable predictora (valor de salida). El número total de características del algoritmo es igual a la entrada, por lo que es posible seleccionar el número de características que se deseen. Un parámetro a definir en este algoritmo es el número  $k$  de vecinos a evaluar, en donde no hay regla alguna para seleccionar este parámetro.

En este trabajo se implementó RReliefF usando el comando **relieff** de Matlab V2019b, variando los k-vecinos de 5 a 40 y se probó con 500, 200 y 500 características de entrada para la ANN multicapa. La figura 4.5 muestra el flujo de los datos, empezando por la región espectral que sirve de entrada al algoritmo de RReliefF, en el cual se modificaron los parámetros de número de características y k-vecinos; las características seleccionadas ingresan a la red SOM mediante una validación cruzada de 3 particiones; lo obtenido de este paso se usa como la entrada a la ANN para obtener el porcentaje de HbA1c o la concentración de glucosa.

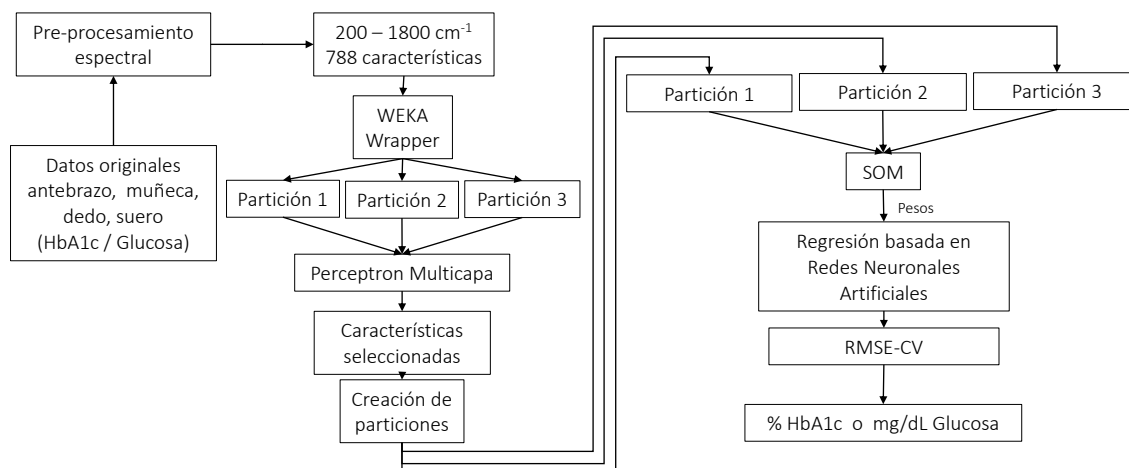


**Figura 4.5:** Sistema de medición utilizando RReliefF.

#### 4.1.5. Selección de características usando Wrapper

El método Wrapper se implementó a través del software WEKA, usando una validación cruzada de 3 particiones y dos algoritmos de búsqueda, al igual que se hizo en CFS. En este caso, se eligió un método de regresión basado en una red neuronal de perceptrón multicapa, en el que se evaluaron las diferentes combinaciones de características seleccionando las características que se utilizaron al menos en una partición. Cabe mencionar que este método de selección de características es computacionalmente costoso. La Figura 4.6 muestra el diagrama descriptivo de esta implementación, en donde se observa la región espectral seleccionada como entrada al algoritmo Wrapper usando validación cruzada y el algoritmo de inducción seleccionado; posteriormente las características seleccionadas ingresan a la red SOM, en donde se generan los pesos correspondientes a cada entrada, siendo éstos los datos que ingresan

a la ANN multicapa.



**Figura 4.6:** Sistema de medición utilizando Wrapper.

## 4.2. Otros métodos utilizados

Para implementar SVM y LR, se utilizó la caja de herramientas Matlab 2019b de “Regression Learner” y se aplicó una validación cruzada de 3 particiones. En el caso de SVM se implementaron todos los kernels de la caja de herramientas (polinomiales y gaussianos). La entrada al regresor fue el conjunto de características en las que los resultados de la ANN fueron los mejores. Para iPLS se utilizó el programa generado por [105] con 10 intervalos espectrales para el análisis; se inspeccionó visualmente que estas particiones no cortaran a la mitad algún pico relevante.

## 4.3. Clasificación

La estimación aproximada de la cantidad de HbA<sub>1c</sub> a través de un modelo de clasificación se realizó de la siguiente manera: primero, para generar el conjunto de entrenamiento de la ANN, los voluntarios se clasificaron según su porcentaje de HbA<sub>1c</sub>, considerando la siguiente distribución la cual es reportada por organizaciones tales como la organización mundial de la salud (OMS) [21]:

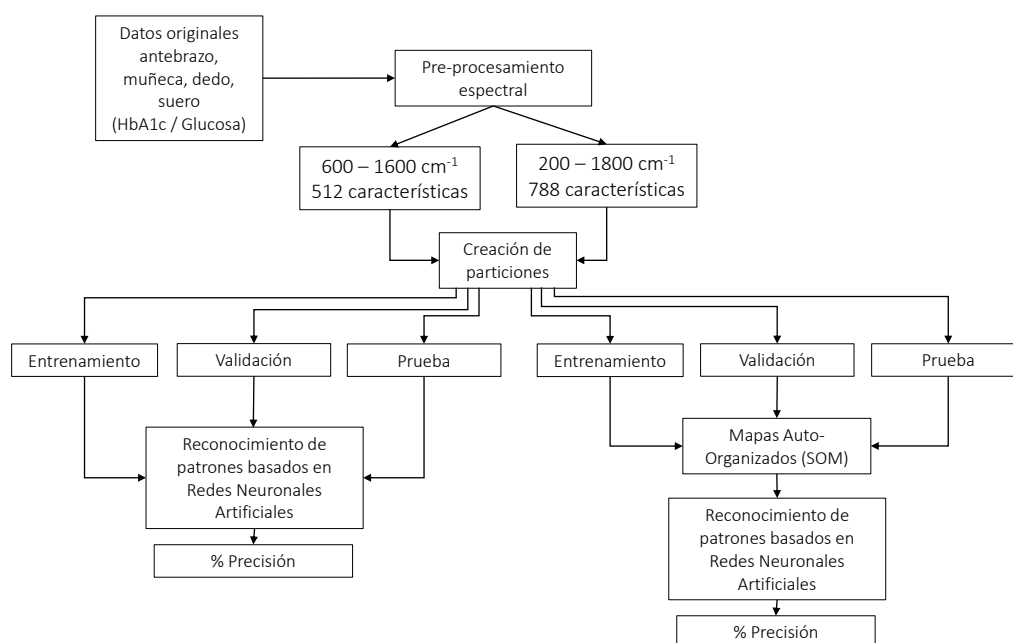
- menor que 5.7 % → No diabéticos
- de 5.7 % a 6.5 % → Pre-diabéticos
- mayor que 6.5 → Diabéticos

Por lo anterior, puede verse que se consideraron tres clases, es decir, es un problema multi-clase. Acorde con la base de datos, 4 voluntarios son no diabéticos, 10 voluntarios pre-diabéticos y 32 voluntarios diabéticos. Se usó la caja de herramientas de reconocimiento de patrones (NPR, por sus siglas en inglés) de Matlab 2019b [146]; el conjunto de datos se dividió en tres conjuntos, entrenamiento 75 %, validación 15 % y prueba 15 %.

En este caso, se probaron los dos rangos espectrales anteriormente mencionados. Los datos de entrada al clasificador fueron pre-procesados, es decir, se eliminó la fluorescencia y fueron suavizados. Por otro lado, considerando el desempeño que tuvo la red SOM en regresión, se realizó



otra prueba usando la red SOM como paso previo a la clasificación. Un diagrama descriptivo se muestra en la figura 4.7, en donde se observan las dos regiones espectrales usadas en este caso, la generación de particiones y el uso de reconocimiento de patrones basado en ANN; además se incluye la implementación de la Red SOM como en los casos de regresión.



**Figura 4.7:** Sistema de medición utilizando un clasificador.

## 4.4. Análisis espectral

A partir de la HbA1c y la glucosa en polvo, se obtuvieron los espectros Raman y sus picos representativos mediante un ajuste gaussiano utilizando un software basado en algoritmos genéticos hecho a mano por uno de los integrantes del laboratorio GIOB de la coordinación de óptica del INAOE [98]. Los picos encontrados se asociaron a su vibración molecular de acuerdo con lo reportado en la literatura.

De igual manera se analizaron los espectros Raman obtenidos del antebrazo, muñeca, dedo de los voluntarios, así como los obtenidos de suero, cada base de datos contiene 36 concentraciones diferentes de HbA1c y 43 concentraciones de glucosa. Se calculó el promedio de cada concentración y se realizó el

mismo ajuste gaussiano que se menciona en el párrafo anterior. Finalmente, se identificaron los picos encontrados en la sustancia pura y los picos encontrados en las mediciones *in vivo* e *in vitro*.

## 4.5. Resumen del capítulo

En este capítulo se presentó una descripción detallada de la implementación de métodos computacionales para la cuantificación de glucosa y HbA<sub>1c</sub>. Cabe resaltar que este trabajo presenta el uso por primera vez de algunos métodos de extracción y selección de características, tales como RReliefF, CFS y Wrapper. La implementación de diferentes métodos de refinamiento y selección de características tiene la finalidad de obtener el mínimo error en la cuantificación no invasiva de HbA<sub>1c</sub> y glucosa. En el siguiente capítulo se muestran los resultados obtenidos en dichas implementaciones.

---

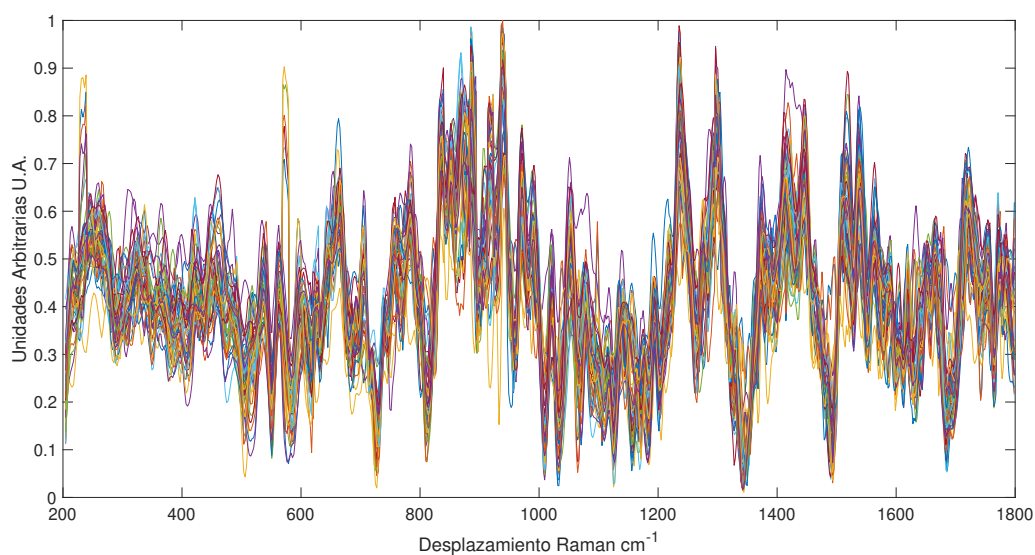
## RESULTADOS Y DISCUSIÓN

---

En este capítulo se presentan los resultados de la implementación de los métodos descritos en el capítulo 4. Estos resultados se agrupan en regresión y clasificación, en la parte de regresión se analizarán los resultados de cada una de las técnicas computacionales, así como la comparativa con otros métodos utilizados en la literatura como lo son SVM, LR e iPLS; en la parte de clasificación se mostrarán los resultados obtenidos del análisis de 3 clases basadas en el porcentaje de HbA<sub>1c</sub> por voluntario. Así mismo se presenta un análisis espectral de las sustancias puras en donde se identifican sus bandas Raman conforme a lo reportado en la literatura y la localización de estas bandas en las mediciones *in vivo* de las diferentes partes del cuerpo analizadas.

### 5.1. Base de datos mediciones *in vivo*

Como anteriormente se mencionó, se realizaron mediciones de espectros Raman en tres diferentes partes del cuerpo a 46 voluntarios; en la figura 5.1 se observa el promedio de 9 mediciones por voluntario para la región de antebrazo ( para dedo, muñeca y suero véase los apéndices A.1, A.2, A.3), cabe mencionar que estos espectros ya están filtrados y normalizados.



**Figura 5.1:** Espectros Raman promedio por voluntario en la región de antebrazo.

En la tabla 5.1 se muestran los valores de HbA<sub>1c</sub>, glucosa, sexo (M masculino, F femenino), edad, peso y estatura de los 46 voluntarios. Sin embargo, el historial clínico de estos voluntarios se desconoce, así mismo el peso y la estatura son aproximados, puesto que son los valores que los mismos voluntarios proporcionaron.

Tabla 5.1: Datos de voluntarios

| ID GIOB | Glucosa (mg/dL) | HbA1c (%) | Sexo | Edad (años) | Estatura (cm) | Peso (kg) |
|---------|-----------------|-----------|------|-------------|---------------|-----------|
| V1      | 112             | 6.3       | M    | 45          | 160           | 73        |
| V2      | 129             | 8.6       | F    | 46          | 154           | 83        |
| V3      | 74              | 5.3       | F    | 48          | 156           | 70        |
| V4      | 140             | 8         | M    | 48          | 170           | 83        |
| V5      | 189             | 10        | M    | 48          | 179           | 79        |
| V6      | 95              | 7.3       | F    | 31          | 160           | 81        |
| V7      | 275             | 14        | M    | 53          | 155           | 48        |
| V8      | 106             | 9.3       | F    | 54          | 160           | 62        |
| V9      | 400             | 13.8      | M    | 53          | 180           | 80        |
| V10     | 195             | 12.2      | F    | 49          | 155           | 59.3      |
| V11     | 125             | 10.2      | F    | 44          | 155           | 67        |
| V12     | 107             | 5.4       | F    | 47          | 150           | 69.5      |
| V13     | 100             | 6.6       | F    | 49          | 156           | 59        |
| V14     | 101             | 6.3       | F    | 47          | 156           | 67        |
| V15     | 72              | 6.2       | F    | 48          | 155           | 70        |
| V16     | 89              | 6.4       | F    | 61          | 150           | 86        |
| V17     | 109             | 6.8       | F    | 54          | 153           | 78        |
| V18     | 99              | 10        | F    | 48          | 152           | 47        |
| V19     | 127             | 10        | M    | 47          | 176           | 125       |
| V20     | 214             | 13.4      | F    | 38          | 170           | 98        |
| V21     | 80              | 8.6       | M    | 59          | 152           | 56        |
| V22     | 56              | 8.6       | M    | 66          | 155           | 56        |
| V23     | 163             | 9         | M    | 48          | 166           | 84.5      |
| V24     | 74              | 6.7       | M    | 55          | 163           | 73        |
| V25     | 104             | 7.4       | F    | 41          | 170           | 84        |
| V26     | 169             | 10.1      | F    | 52          | 165           | 99        |
| V27     | 165             | 11.1      | M    | 56          | 164           | 61        |
| V28     | 115             | 11.8      | F    | 87          | 160           | 74        |
| V29     | 225             | 12.2      | F    | 43          | 160           | 86        |
| V30     | 268             | 12.1      | M    | 45          | 172           | 91        |
| V31     | 226             | 12.7      | F    | 57          | 158           | 45        |
| V32     | 117             | 6.5       | F    | 42          | 145           | 71        |
| V33     | 90              | 5.5       | F    | 40          | 150           | 51        |
| V34     | 290             | 9.8       | F    | 45          | 160           | 76        |
| V35     | 101             | 6.1       | F    | 50          | 145           | 64.5      |
| V36     | 108             | 5.2       | M    | 39          | 177           | 89        |
| V37     | 126             | 8.1       | F    | 52          | 170           | 72.8      |
| V38     | 95              | 6.2       | F    | 58          | 158           | 61        |
| V39     | 156             | 6         | F    | 78          | 142           | 49        |
| V40     | 228             | 11.2      | F    | 46          | 150           | 60        |
| V41     | 153             | 7.8       | F    | 72          | 155           | 90        |
| V42     | 187             | 7.9       | F    | 64          | 148           | 50.3      |
| V43     | 238             | 8.3       | M    | 45          | 162           | 83        |
| V44     | 260             | 10.2      | M    | 47          | 170           | 65        |
| V45     | 145             | 6.4       | M    | 79          | 164           | 68        |
| V46     | 107             | 6.2       | F    | 27          | 160           | 87        |

## 5.2. Regresión

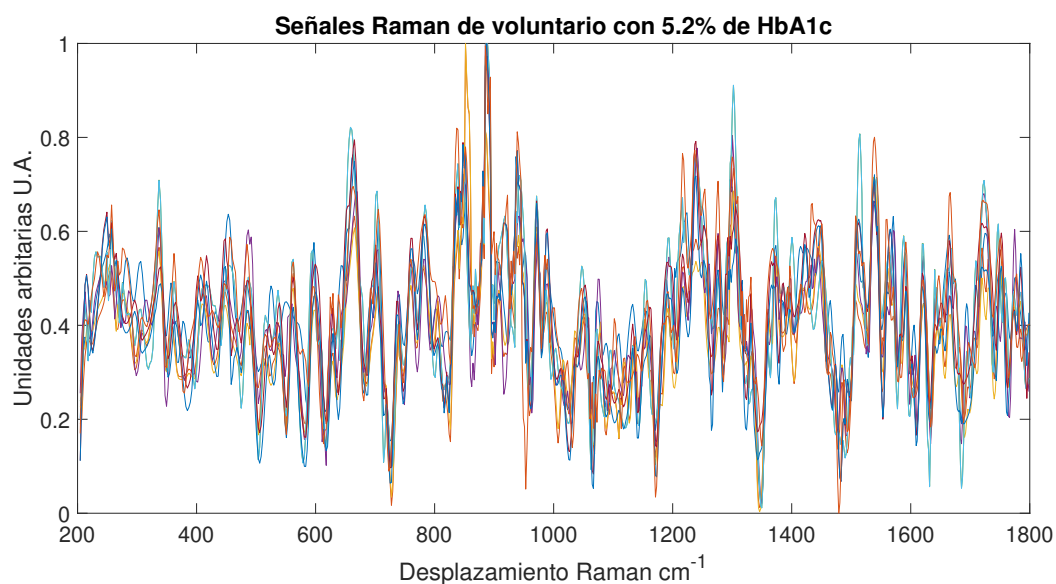
En primera instancia se introdujeron los espectros Raman sin procesamiento a la ANN, sin embargo, los resultados obtenidos no fueron favorable puesto que el RMSE-CV fue mayor al 1 % de HbA<sub>1c</sub>. Es importante recordar que la meta de este trabajo es la cuantificación de concentraciones tanto de glucosa como HbA<sub>1c</sub>; sin embargo, para que este método sea funcional y pueda ser comparable con métodos ya establecidos es necesario disminuir el error lo más que sea posible, por lo que se propusieron métodos alternativos para lograr este objetivo. A continuación, se presentan los resultados obtenidos de cada uno de los métodos empleados. Cabe resaltar que estos métodos se propusieron a partir de que los resultados obtenidos de ingresar el espectro Raman directamente a la ANN no fueron favorables, los resultados se muestran en la tabla 5.2.

**Tabla 5.2:** Resultado de la ANN con datos sin procesar.

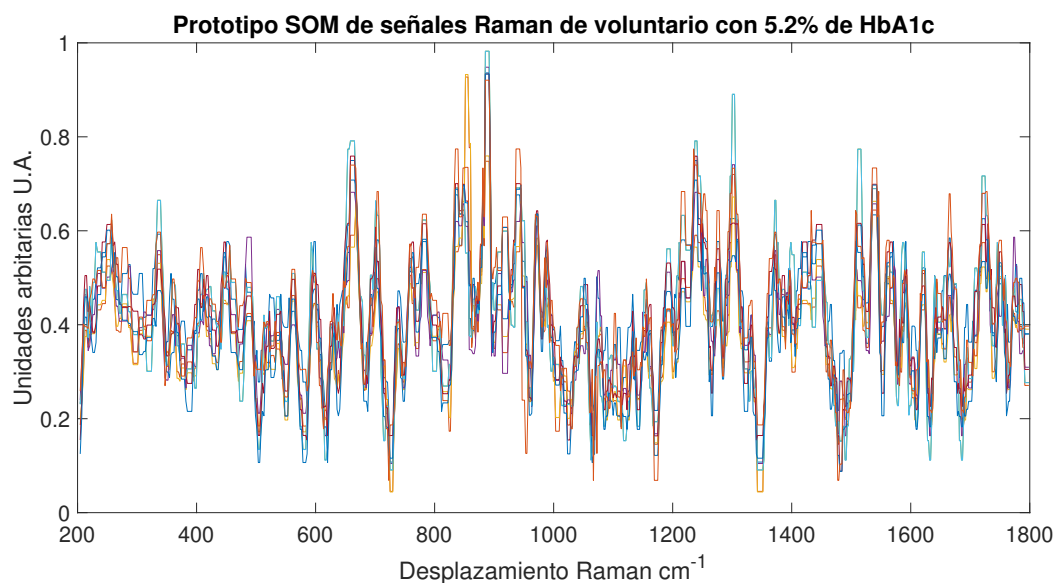
| Región corporal | RMSE-CV               |                 |
|-----------------|-----------------------|-----------------|
|                 | HbA <sub>1c</sub> (%) | Glucosa (mg/dL) |
| Antebrazo       | 1.03 ± 0.09           | 60.32 ± 5.27    |
| Muñeca          | 1.08 ± 0.03           | 63.12 ± 6.89    |
| Dedo            | 2.01 ± 0.09           | 68.30 ± 6.57    |
| Suero           | NA                    | 23.50 ± 5.72    |

### 5.2.1. Mapas auto-organizados

En la figura 5.2 se muestran los espectros Raman de 9 mediciones realizadas en la muñeca de un voluntario con 5.2 % de HbA<sub>1c</sub>, estos espectros así como los de los otros 45 voluntarios fueron ingresados a la red SOM como anteriormente se mencionó, lo que se obtuvo como resultado es lo que se muestra en la figura 5.3, donde se observan los prototipos generados a partir de las señales de entrada, estas señales fueron la entrada a la red neuronal.



**Figura 5.2:** Espectro Raman de un voluntario con 5.2 % de HbA1c.



**Figura 5.3:** Prototipo Red SOM del espectro Raman del voluntario con 5.2 % de HbA1c.

En la tabla 5.3 se observan los resultados obtenidos en cada una de las zonas corporales analizadas usando las dos regiones espectrales (200 - 1800  $\text{cm}^{-1}$  y 600 - 1600  $\text{cm}^{-1}$ ), como bien se mencionó anteriormente el número de características que ingresan a la red neuronal son el mismo que ingresó a la red SOM, al incrementar el número de características el tiempo de cómputo

incrementa por lo que para 788 características el tiempo computacional fue mayor que para 512, sin embargo de los resultados obtenidos de las tres partes del cuerpo analizadas, se observa que en este caso el mejor resultado se obtiene para las 788 características, es decir, todo el rango espectral y para el área de la muñeca, con un error de  $0.68 \pm 0.01 \%$ , es importante recordar que el rango de medición de HbA<sub>1c</sub> es de 5.2 % a 14 %.

**Tabla 5.3:** Resultados de RMSE-CV para concentraciones de HbA<sub>1c</sub> usando Red SOM.

| Rango espectral 200-1800 $\text{cm}^{-1}$ (788 características)   |                 |
|---|-----------------|
| Región corporal   | RMSE-CV (%)     |
| Antebrazo   | $0.70 \pm 0.01$ |
| Muñeca  | $0.68 \pm 0.01$ |
| Dedo  | $1.95 \pm 0.17$ |
| Rango espectral 600 - 1600 $\text{cm}^{-1}$ (512 características) |                 |
| Antebrazo   | $1.41 \pm 0.04$ |
| Muñeca  | $0.85 \pm 0.02$ |
| Dedo  | $1.90 \pm 0.06$ |

Por otro lado, en el caso de las mediciones de glucosa los resultados pueden dividirse en: mediciones *in vivo* y mediciones *in vitro* (realizadas al suero); en la tabla 5.4 se observa el mínimo RMSE-CV obtenido al utilizar antebrazo en un rango de 600 a 1600  $\text{cm}^{-1}$  con un error de  $56.31 \pm 4.28$ ; mg/dL; en el caso del suero el error fue de  $22.49 \pm 7.81$  mg/dL para el rango de 200 a 1800  $\text{cm}^{-1}$ . Sin embargo, si se hace una comparación entre el error obtenido con HbA<sub>1c</sub>, glucosa y sus respectivas concentraciones el error es mucho mayor en el caso de la glucosa esto puede deberse a que el rango de valores para glucosa es de 56 a 400 mg/dL, un rango de valores mayor comparado al de HbA<sub>1c</sub> o que este tipo de procesamiento no genera una representación de datos que favorezca a la predicción de glucosa.



**Tabla 5.4:** Resultados de RMSE-CV para glucosa usando Red SOM.

| Rango espectral 200-1800 $\text{cm}^{-1}$ (788 características)   |                   |
|---|-------------------|
| Región corporal   | RMSE-CV (mg/dL)   |
| Antebrazo   | $58.43 \pm 5.95$  |
| Muñeca  | $56.42 \pm 1.79$  |
| Dedo  | $59.25 \pm 1.56$  |
| Suero   | $22.49 \pm 7.81$  |
| Rango espectral 600 - 1600 $\text{cm}^{-1}$ (512 características) |                   |
| Antebrazo   | $56.31 \pm 4.28$  |
| Muñeca  | $58.22 \pm 1.03$  |
| Dedo  | $56.65 \pm 8.99$  |
| Suero   | $28.53 \pm 13.27$ |

### 5.2.2. Análisis de Componentes Principales

Esta técnica es muy usada en estudios quimiométricos (aplicación de métodos matemáticos o estadísticos sobre datos químicos [153]), con la finalidad de tener una representación de los en una dimensionalidad menor, es decir, menos datos como entrada a la ANN; los resultados obtenidos se muestran en la tabla 5.5 donde se observa que el RMSE-CV incremento considerablemente al usar PCA, con un mínimo de error de  $1.65 \pm 0.14 \%$  para el caso del dedo en el rango de 200 a  $1800 \text{ cm}^{-1}$  con una representación de 100 PC y 99 % de varianza acumulada; comparándolo con lo obtenido con la red SOM en este mismo caso se observa una disminución de error de  $0.25 \%$  y una reducción de 412 características; aunque para las demás partes del cuerpo este análisis no obtuvo una disminución en el error.

El caso de la cuantificación de glucosa *in vivo* se presenta en la tabla 5.6, en donde se observa un incremento considerable en el error comparándolo con lo obtenido anteriormente con la red SOM, de igual manera que en el caso de cuantificación de HbA<sub>1c</sub> en la región del dedo se presentó una disminución del error del  $9.5 \%$ , aunado a esto la variación entre perdición por partición de datos también disminuyo de  $8.99 \%$  a  $0.89 \%$ . Cabe resaltar que estos resultados son solamente aplicando PCA y ANN.

**Tabla 5.5:** Resultados de RMSE-CV para HbA<sub>1c</sub> usando PCA

| Rango espectral 200 - 1800 cm <sup>-1</sup> (788 características) |      |                      |             |
|---|------|----------------------|-------------|
| Región corporal   | # PC | % Varianza acumulada | RMSE-CV (%) |
| Antebrazo   | 100  | 95                   | 2.30 ± 0.26 |
| Muñeca  | 100  | 96                   | 2.25 ± 0.16 |
| Dedo  | 100  | 99                   | 1.65 ± 0.14 |
| Rango espectral 600 - 1600 cm <sup>-1</sup> (512 características) |      |                      |             |
| Antebrazo   | 100  | 97                   | 2.29 ± 0.52 |
| Muñeca  | 100  | 97                   | 2.09 ± 0.16 |
| Dedo  | 100  | 99                   | 1.70 ± 0.06 |

**Tabla 5.6:** Resultados de RMSE-CV para glucosa usando PCA.

| Rango espectral 200 - 1800 cm <sup>-1</sup> |      |                      |                 |
|---|------|----------------------|-----------------|
| Región corporal                             | # PC | % Varianza acumulada | RMSE-CV (mg/dL) |
| Antebrazo                                   | 100  | 95                   | 73.91 ± 5.35    |
| Muñeca                                      | 100  | 96                   | 74.66 ± 5.46    |
| Dedo  | 100  | 99                   | 51.50 ± 3.60    |
| Suero                                       | 100  | 84                   | 48.02 ± 3.29    |
| Rango espectral 600 - 1600 cm <sup>-1</sup> |      |                      |                 |
| Antebrazo                                   | 100  | 97                   | 73.22 ± 6.96    |
| Muñeca                                      | 100  | 98                   | 76.50 ± 4.93    |
| Dedo  | 100  | 99                   | 47.15 ± 0.89    |
| Suero                                       | 100  | 98                   | 49.44 ± 4.16    |

Considerando la disminución del error obtenida mediante el uso de la red SOM se decidió implementar en conjunto con PCA, los resultados se muestran en las siguientes tablas 5.7, 5.8. La tabla 5.7 muestra los resultados obtenidos de la cuantificación de HbA<sub>1c</sub> en donde se observa que existe una disminución del error para antebrazo y muñeca, sin embargo, el error para el dedo incrementa. También cabe resaltar que en este método se probaron las dos regiones espectrales mencionadas anteriormente, puesto que como se ha mencionado PCA realiza una transformación de los datos y esta transformación se ve afectada por este factor.

En el caso de la glucosa se puede observar para todos los casos se obtuvo un

**Tabla 5.7:** Resultados de RMSE-CV para HbA<sub>1c</sub> usando PCA-SOM.

| Rango espectral 200 - 1800 cm <sup>-1</sup> |      |                      |             |
|---|------|----------------------|-------------|
| Región corporal                             | # PC | % Varianza acumulada | RMSE-CV (%) |
| Antebrazo                                   | 100  | 95                   | 1.71 ± 0.14 |
| Muñeca                                      | 100  | 96                   | 1.78 ± 0.14 |
| Dedo  | 100  | 81                   | 1.81 ± 0.10 |
| Rango espectral 600 - 1600 cm <sup>-1</sup> |      |                      |             |
| Antebrazo                                   | 100  | 97                   | 1.73 ± 0.15 |
| Muñeca                                      | 100  | 97                   | 1.86 ± 0.12 |
| Dedo  | 100  | 81                   | 1.86 ± 0.13 |

menor error. En la tabla 5.8 se presentan los resultados numéricos, a partir de estos resultados se infiere que en el caso de glucosa implementar red SOM-PCA produce una mejora en la predicción de concentraciones usando ANN.

**Tabla 5.8:** Resultados de RMSE-CV para glucosa usando PCA-SOM.

| Rango espectral 200 - 1800 cm <sup>-1</sup> (788 características) |      |                      |                 |
|---|------|----------------------|-----------------|
| Región corporal   | # PC | % Varianza acumulada | RMSE-CV (mg/dL) |
| Antebrazo   | 100  | 95                   | 62.24 ± 8.91    |
| Muñeca  | 100  | 96                   | 63.97 ± 2.37    |
| Dedo  | 100  | 91                   | 38.23 ± 4.45    |
| Suero   | 100  | 98                   | 27.37 ± 7.86    |
| Rango espectral 600 - 1600 cm <sup>-1</sup> (512 características) |      |                      |                 |
| Antebrazo   | 100  | 97                   | 67.77 ± 10.28   |
| Muñeca  | 100  | 98                   | 63.51 ± 6.11    |
| Dedo  | 100  | 89                   | 39.39 ± 5.99    |
| Suero   | 100  | 99                   | 30.86 ± 5.99    |

La mejora al aplicar red SOM se puede deber a que al ser el objetivo general de la red SOM encontrar patrones entre los datos, esto ayuda a la generación de prototipos mediante la extracción de los pesos que mejoran en casi la mayoría de casos las señales de entrada, sean estas los espectros completos, PCA u otros. Esto a su vez depende de la región corporal analizada puesto que las señales que ingresan a los métodos son diferentes y esto puede deberse a la

distancia existente entre las capas de piel, músculo, entre otros componentes y la sangre, que es la sustancia que se quiere analizar debido a que contiene los biomarcadores de interés. Otro parámetro que influye en estos resultados es el valor de salida esperado, puesto como podemos observar en el caso del dedo para la determinación de HbA<sub>1c</sub> la implementación de la red SOM a PCA no reduce el error comparado con PCA.

### 5.2.3. Selección de características basada en correlación

Los resultados obtenidos con el uso de CFS se muestran en la tabla 5.9, en este caso al usar CFS-SOM el error disminuye para antebrazo y muñeca, sin embargo, incrementa para el dedo; así mismo de los dos métodos de búsqueda “BestFirst” funciona mejor para antebrazo y muñeca; “GreedyStepwise” presentó un mejor resultado para el dedo. En este caso podemos observar que, a pesar de tener un número menor de características que en PCA-SOM, los resultados en estas dos partes del cuerpo mejoraron a comparación de ser representados por los 100 PC. Esto puede deberse a la complejidad espectral de la señal haciendo que la transformación a las componentes principales no favorezca a los resultados, en cambio de conservar los datos originales y seleccionarlos. Sin embargo, para el dedo se observa que con el algoritmo de búsqueda GreedyStepwise y 16 características seleccionadas obtuvo un mejor resultado que BestFirst y 28 características esto quiere decir que, en este específico caso, a pesar de ser más características no están representando de manera adecuada a los datos. Cabe resaltar que estas características seleccionadas no necesariamente corresponden a los picos de la sustancia analizada.

En el caso de glucosa la tabla 5.10, se observa que para las mediciones *in vitro* del suero esta representación de los datos y usando el algoritmo de búsqueda “GreedyStepwise”, presenta una mejora comparada con PCA-SOM (al igual que en el caso de la HbA<sub>1c</sub>) sin embargo, la red SOM por si sola en el rango de 200 a 1800 cm<sup>-1</sup> sigue obteniendo un mejor RMSE-CV de 22.49 ± 7.81 mg/dL, aunque el número de características disminuye de 788 a 30 y solo presenta un incremento de error del 0.43 %. En este caso al ser una aplicación médica se requiere que el error sea lo más cercano a 0 posible.

Por otro lado, en el caso de las mediciones *in vivo*, solo para el antebrazo y

**Tabla 5.9:** Resultados de RMSE-CV para HbA<sub>1c</sub> usando CFS-SOM.

| Algoritmo de búsqueda: "BestFirst"      |                   |             |
|---|-------------------|-------------|
| Región corporal                         | # Características | RMSE-CV (%) |
| Antebrazo                               | 27                | 1.60 ± 0.05 |
| Muñeca                                  | 28                | 1.52 ± 0.07 |
| Dedo                                    | 28                | 2.07 ± 0.24 |
| Algoritmo de búsqueda: "GreedyStepwise" |                   |             |
| Antebrazo                               | 21                | 1.66 ± 0.10 |
| Muñeca                                  | 24                | 1.56 ± 0.09 |
| Dedo                                    | 16                | 1.94 ± 0.01 |

muñeca se observa una mejora, aunque para el dedo en error se incrementa. Esto puede deberse al número de elementos en la base de datos; es importante recordar que para el caso de antebrazo y muñeca se adquirieron 9 espectros por voluntario y en el caso de dedo solamente 3 por voluntario.

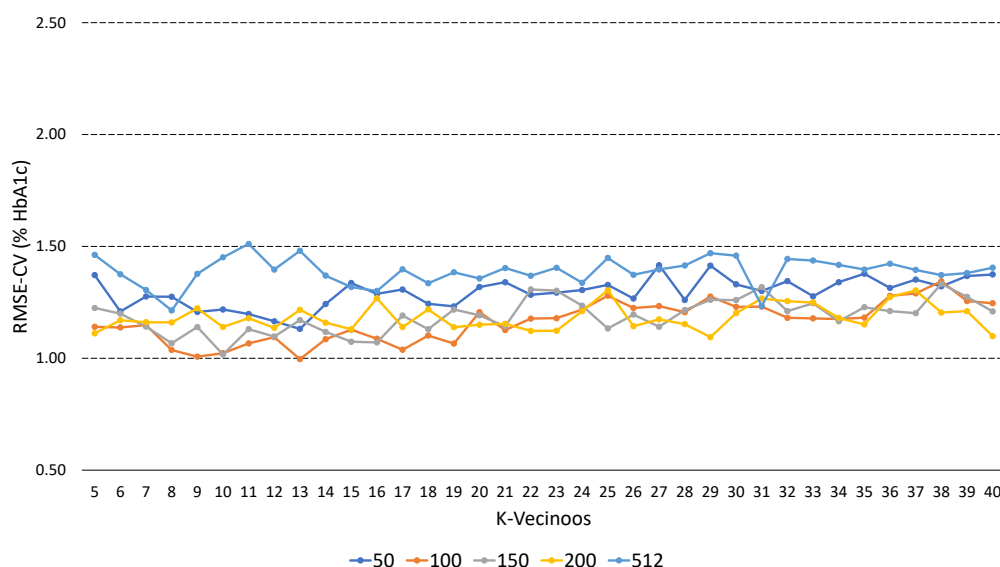
**Tabla 5.10:** Resultados de RMSE-CV para glucosa usando CFS-SOM.

| Algoritmo de búsqueda: "BestFirst"      |                   |                 |
|---|-------------------|-----------------|
| Región corporal                         | # Características | RMSE-CV (mg/dL) |
| Antebrazo                               | 31                | 46.23 ± 2.76    |
| Muñeca                                  | 21                | 55.44 ± 2.21    |
| Dedo                                    | 21                | 67.65 ± 7.76    |
| Suero                                   | 16                | 41.39 ± 7.43    |
| Algoritmo de búsqueda: "GreedyStepwise" |                   |                 |
| Antebrazo                               | 21                | 48.72 ± 5.39    |
| Muñeca                                  | 13                | 57.60 ± 1.94    |
| Dedo                                    | 17                | 65.82 ± 9.63    |
| Suero                                   | 30                | 22.92 ± 5.68    |

#### 5.2.4. RReliefF

Con este algoritmo se probaron inicialmente 10, 20 y 30 vecinos con 100, 200 y 300 características, sin embargo se observó que eran muy pocas pruebas

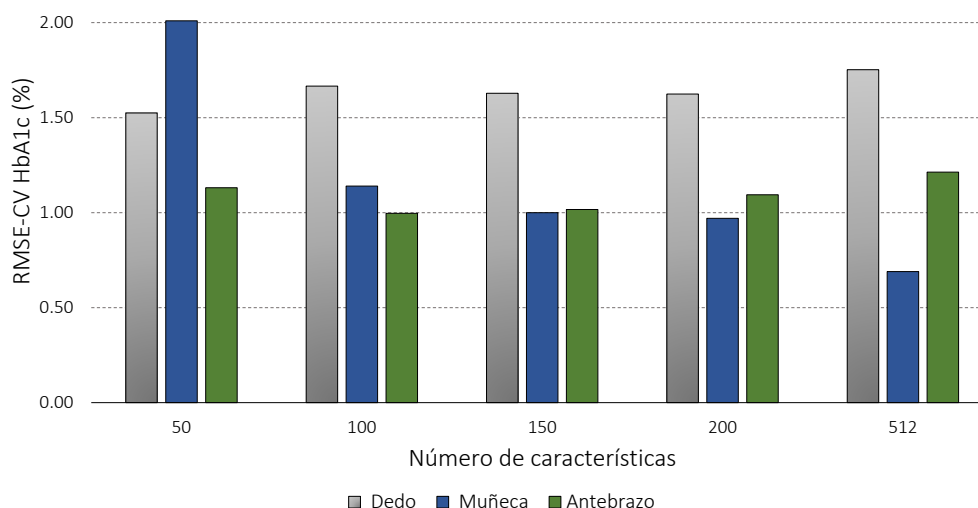
para tomar una decisión contundente, así que se decidió hacer un barrido entre el número de k-vecinos y el número de características, esto permitió observar que la variación de vecinos en su mayoría no presenta fluctuaciones notorias en los resultados, como se observa en la figura 5.4, en donde 50, 100, 150, 200 y 512 son las características utilizadas para predicción de HbA1c en antebrazo, este comportamiento se presenta de manera similar para las otras dos partes del cuerpo ( véase apéndices A.4,A.5) .



**Figura 5.4:** Ejemplo variación de K-Vecinos por número de características en la región del antebrazo.

Es interesante notar que en estas pruebas el error mínimo se obtuvo en un determinado número de características y después incrementó como se observa en la figura 5.5.

El primer experimento realizado fue con la región de muñeca y se observó que conforme se incrementaba el número de características el error disminuía, sin embargo al analizar los demás casos se observó que este comportamiento era distinto, por ejemplo, para el dedo el mejor resultado se obtuvo para 50 características, lo que concuerda con lo que observamos en PCA en donde el dedo presentó mejor desempeño con un número menor de características, sin embargo estas características tiene que ser representativas del conjunto de datos como se observó en CFS, donde las 28 características no fueron representativas obteniendo un error mayor. En el caso del antebrazo el mínimo



**Figura 5.5:** RMSE-CV por parte del cuerpo y número de características usando RReliefF.

error se obtuvo con 100 características, otro resultado importante de resaltar para la cuantificación de HbA1c es que no es necesario probar más allá de los 26 vecinos para este caso, puesto que los mejores resultados se dieron entre 26 y 13 vecinos, como se observa en la tabla 5.11. Cabe resaltar que estas características seleccionadas no necesariamente corresponden a picos de la sustancia analizada.

**Tabla 5.11:** Resultados de RMSE-CV para HbA1c usando RReliefF-SOM.

| Rango espectral 200 - 1800 $\text{cm}^{-1}$ |                   |           |                 |
|---|-------------------|-----------|-----------------|
| Región corporal                             | # características | # Vecinos | RMSE-CV (%)     |
| Antebrazo                                   | 100               | 13        | $1.00 \pm 0.08$ |
| Muñeca                                      | 512               | 26        | $0.69 \pm 0.07$ |
| Dedo  | 50                | 23        | $1.52 \pm 0.41$ |

Los resultados para el caso de glucosa se presentan en la tabla 5.12 donde se puede observar que tanto para muñeca, antebrazo y suero se presentó el mejor de los resultados, con 200 características para el antebrazo y 150 para muñeca, dedo y suero. Sin embargo, en este caso el número de vecinos para dedo se alejó de la media de vecinos de las otras partes del cuerpo, aunque para el

dedo este resultado no es el mejor de los casos como veremos más adelante. Cabe mencionar que el comportamiento entre K-vecinos y características fue similar que en el caso de HbA<sub>1c</sub> (véase apéndice A.6).

**Tabla 5.12:** Resultados de RMSE-CV para glucosa usando RReliefF-SOM.

| Rango espectral 200 - 1800 cm <sup>-1</sup> |                   |           |                 |
|---|-------------------|-----------|-----------------|
| Región corporal                             | # características | # Vecinos | RMSE-CV (mg/dL) |
| Antebrazo                                   | 200               | 9         | 30.12 ± 0.53    |
| Muñeca                                      | 150               | 7         | 41.44 ± 4.68    |
| Dedo  | 150               | 38        | 49.67 ± 1.66    |
| Suero                                       | 150               | 13        | 11.86 ± 5.45    |

### 5.2.5. Wrapper

En el caso del método *Wrapper* es importante considerar el tiempo de cómputo puesto que puede llegar a tardar días solamente en la selección de características esto debido al algoritmo de aprendizaje de inducción seleccionado, puesto que este algoritmo se sugiere sea el que se utilizará posteriormente, en este caso se usó ANN como método para regresión, por lo que en Weka se usó perceptrón multicapa, cabe mencionar que se probó con otro algoritmo y la selección de características se realizó en cuestión de minutos. Otro punto importante a mencionar es que los parámetros dentro de Weka no se modificaron, probablemente si se modificaran los resultados cambien y podríamos llegar a tener un mejor resultado.

De la tabla 5.13 se observa que el método *Wrapper* representa con un menor número de características la señal que CFS obteniendo en general mejores resultados que PCA y CFS. Sin embargo, esta representación de los datos no es suficiente para generar un menor error que RReliefF, el cual fue el mejor resultado que obtuvimos para HbA<sub>1c</sub> y glucosa. Como anteriormente se mencionó las características seleccionadas no necesariamente corresponden a picos de la sustancia analizada.

Otra particularidad que es posible observar en el uso de búsquedas inteligentes es que el algoritmo de GreedyStepwise selecciona menos características que BestFirst pero sus resultados empeoran con esa disminución



y selección, por lo tanto no representa de manera adecuada la señal para la cuantificación de HbA<sub>1c</sub>. Los resultados para glucosa se observan en la tabla

**Tabla 5.13:** Resultados de RMSE-CV para HbA<sub>1c</sub> usando Wrapper-SOM.

| Algoritmo de búsqueda: “ Best First”    |                   |             |
|---|-------------------|-------------|
| Región corporal                         | # características | RMSE-CV (%) |
| Antebrazo                               | 21                | 1.47 ± 0.01 |
| Muñeca                                  | 24                | 1.52 ± 0.03 |
| Dedo                                    | 15                | 1.89 ± 0.22 |
| Algoritmo de búsqueda: “GreedyStepwise” |                   |             |
| Antebrazo                               | 10                | 1.62 ± 0.08 |
| Muñeca                                  | 12                | 1.51 ± 0.05 |
| Dedo                                    | 13                | 1.94 ± 0.13 |

5.14, en este caso solo se tuvo una ligera mejora en el caso de muñeca y dedo con el algoritmo de búsqueda “Best first”, comparándolo con CFS, sin embargo, los errores son mayores a los obtenidos con otros métodos utilizados.

**Tabla 5.14:** Resultados de RMSE-CV para glucosa usando Wrapper-SOM.

| Algoritmo de búsqueda: “ Best First”    |                   |                 |
|---|-------------------|-----------------|
| Región corporal                         | # características | RMSE-CV (mg/dL) |
| Antebrazo                               | 24                | 54.83 ± 6.45    |
| Muñeca                                  | 26                | 53.71 ± 8.59    |
| Dedo                                    | 26                | 62.42 ± 8.38    |
| Suero                                   | 10                | 38.21 ± 4.41    |
| Algoritmo de búsqueda: “GreedyStepwise” |                   |                 |
| Antebrazo                               | 6                 | 64.72 ± 2.61    |
| Muñeca                                  | 9                 | 60.39 ± 2.75    |
| Dedo                                    | 9                 | 67.25 ± 5.14    |
| Suero                                   | 16                | 42.43 ± 1.88    |

A continuación se presenta una tabla comparativa de los tiempos de cómputo de los métodos implementados para el caso de HbA<sub>1c</sub>. Cabe resaltar que para la predicción de los porcentajes de HbA<sub>1c</sub> usando espectros sin procesar en la región 1 (200 a los 1800 cm<sup>-1</sup>) solamente se probó con 10 nodos escondidos en la capa oculta, debido a que la computadora con la que se analizaron los datos (ASUS S501U, procesador i5-8250U, CPU 1.60 GHz, tarjeta gráfica NVIDIA GFORCE 930MX, 8Gb de RAM) no permitió el análisis de más nodos. En esta tabla (5.15) se observa que el método que necesita más tiempo de cómputo es Wrapper; a pesar de ser pocas características las que se seleccionan para definir el problema, en solo la selección de éstas es donde mayor tiempo computacional consume, debido a que el algoritmo de inducción que se utilizó en ese método fue perceptrón multicapa. Por otro lado, el método PCA es el que más rápido se desempeña, puesto que para esto se seleccionaron solo 100 características, aunque no el método que mejor funciona. Es importante también notar que en función de que el número de características incrementa, el tiempo computación en la red neuronal de MATLAB incrementa, por eso se puede observar que tanto en espectros sin procesar en la región 1 (R<sub>1</sub> = 788 características) y 2 (R<sub>2</sub> = 512 características), como en red SOM región 2 (R<sub>2</sub> = 512 características) y RReliefF (50, 100, 150, 200 y 512 características) tienen tiempos de cómputo alto comparado con PCA y CFS.

### 5.3. Otros métodos utilizados

Considerando los resultados obtenidos con redes neuronales artificiales y los diferentes métodos de selección de características y red SOM, se realizó una comparativa con otros métodos usados en la literatura para cálculo de concentraciones de sustancias, tales como SVM, LR y PLS [50, 46, 105, 154, 61]. En la tabla 5.16 se muestran los resultados de utilizar los mejores casos obtenidos en la red neuronal, es decir, para antebrazo el mejor caso fue usando la red SOM en la región espectral de 200 a 1800 cm<sup>-1</sup>, los pesos generados que sirvieron de entrada a la ANN, son la entrada para SVM y LR, de igual manera se hizo para las otras dos regiones corporales, para muñeca el mejor caso fue aplicando RReliefF-SOM y 512 características, cabe mencionar que estas características no corresponden a la región espectral 2, puesto que son 512

**Tabla 5.15:** Complejidad computación, basada en el tiempo de cómputo.

| Procesamiento de datos     | Tiempo (min) |            |            |
|----------------------------|--------------|------------|------------|
|                            | Muñeca       | Antebrazo  | Dedo       |
| Sin procesar R1 (10 nodos) | 14.581       | 10.365     | 14.793     |
| Sin procesar R2            | 3191.016     | 2775.979   | 611.293    |
| SOM R2                     | 3207.682     | 2802.196   | 632.308    |
| PCA R1                     | 3.428        | 8.434      | 4.928      |
| PCA R2                     | 1.851        | 2.890      | 2.157      |
| PCA-SOM R1                 | 3.422        | 7.475      | 4.828      |
| PCA-SOM R2                 | 1.953        | 9.988      | 2.328      |
| CFS-BF                     | 34.514       | 30.297     | 25.039     |
| CFS-GSW                    | 32.094       | 29.240     | 23.191     |
| Wrapper-BF                 | 345629.564   | 345628.267 | 302420.181 |
| Wrapper-GSW                | 345632.327   | 345626.085 | 302431.554 |
| RReliefF                   | 4013.958     | 3927.955   | 1651.458   |

características que RReliefF eligió; y para dedo el mejor de los casos fue usando 50 características seleccionadas con RReliefF-SOM. En los resultados (Tabla 5.16) se observa que el error es mayor a lo obtenido con ANN para HbA<sub>1c</sub> ( $0.69 \pm 0.07\%$ ), y comparando estos dos métodos entre si SVM tiene mejor desempeño, esto se puede deber a que se utilizaron diferentes kernels para realizar la regresión, es decir, kernels que no son necesariamente lineales, esto quiere decir que, en este tipo de problema, las regresiones lineales no presentan un buen desempeño para cuantificar HbA<sub>1c</sub>.

En el caso de glucosa los resultados se observan en la tabla 5.17, en donde el caso de antebrazo y suero para LR y SVM se aproximan mucho a lo obtenido con la ANN. Sin embargo, nuestra propuesta sigue siendo mejor, recordando que para suero se obtuvo un error de  $11.86 \pm 5.45$  mg/dL y para mediciones *in vivo*  $30.12 \pm 0.53$  mg/dL.

Otro método utilizado fue iPLS [105], la tabla 5.18 muestra los resultados con los intervalos que obtuvieron el mejor resultado, como se puede observar los errores obtenidos con mínimos cuadrados parciales son mayores a los obtenidos con SVM; sin embargo, al implementar ANN a esas regiones espectrales se observa que tiene un peor desempeño que PLS para la región de antebrazo y muñeca, esto quiere decir que esta específica representación de datos no es

**Tabla 5.16:** Resultados de RMSE-CV para HbA<sub>1c</sub> comparación con SVM y LR.

| Maquina de Soporte Vectorial |                    |                 |
|------------------------------|--------------------|-----------------|
| Región corporal              | # características  | RMSE-CV (%)     |
| Antebrazo                    | 788 (SOM)          | $1.53 \pm 0.12$ |
| Muñeca                       | 512 (RReliefF-SOM) | $1.40 \pm 0.17$ |
| Dedo                         | 50 (RReliefF-SOM)  | $1.58 \pm 0.12$ |
| Regresión lineal             |                    |                 |
| Antebrazo                    | 788 (SOM)          | $1.76 \pm 0.28$ |
| Muñeca                       | 512 (RReliefF-SOM) | $2.34 \pm 1.69$ |
| Dedo                         | 50 (RReliefF-SOM)  | $2.17 \pm 1.12$ |

**Tabla 5.17:** Resultados de RMSE-CV para glucosa comparación con SVM y LR.

| Maquina de Soporte Vectorial |                    |                  |
|------------------------------|--------------------|------------------|
| Región corporal              | # características  | RMSE-CV (mg/dL)  |
| Antebrazo                    | 200 (RReliefF-SOM) | $30.50 \pm 3.73$ |
| Muñeca                       | 150 (RReliefF-SOM) | $45.54 \pm 7.25$ |
| Dedo                         | 100 (PCA-SOM)      | $54.90 \pm 8.61$ |
| Suero                        | 150 (RReliefF-SOM) | $14.09 \pm 4.35$ |
| Regresión lineal             |                    |                  |
| Antebrazo                    | 200 (RReliefF-SOM) | $30.56 \pm 3.86$ |
| Muñeca                       | 150 (RReliefF-SOM) | $79.63 \pm 7.96$ |
| Dedo                         | 100 (PCA-SOM)      | $59.54 \pm 9.67$ |
| Suero                        | 150 (RReliefF-SOM) | $12.37 \pm 4.61$ |

suficiente para definir de manera adecuada este problema en una ANN.

**Tabla 5.18:** Resultados de RMSE-CV para HbA<sub>1c</sub> comparación iPLS y ANN.

| Región corporal | Intervalo (cm <sup>-1</sup> ) | iPLS (RMSE-CV %) | ANN (RMSE-CV %) |
|-----------------|-------------------------------|------------------|-----------------|
| Antebrazo       | 770.92 - 934.72               | 1.85             | $2.22 \pm 0.24$ |
| Muñeca          | 770.92 - 934.72               | 1.94             | $2.10 \pm 0.10$ |
| Dedo            | 406.82 - 592.61               | 1.96             | $1.75 \pm 0.08$ |

La tabla 5.19 muestra los resultados de implementar iPLS a los espectros de glucosa, de igual manera que en el caso de HbA<sub>1c</sub>, se observa que para muñeca y antebrazo la ANN presenta un desempeño peor a PLS, aunque para dedo y suero las regiones seleccionadas tienen un menor error.

**Tabla 5.19:** Resultados de RMSE-CV para glucosa comparación iPLS y ANN.

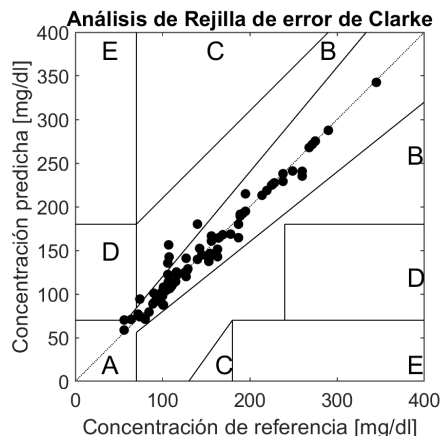
| Región corporal | Intervalo (cm <sup>-1</sup> ) | iPLS (RMSE-CV mg/dL) | ANN (RMSE-CV mg/dL) |
|-----------------|-------------------------------|----------------------|---------------------|
| Antebrazo       | 1389.6-1528.6                 | 65.73                | 73.01 ± 7.45        |
| Muñeca          | 770.92-934.72                 | 64.55                | 73.73 ± 3.63        |
| Dedo            | 936.76-1092.2                 | 67.42                | 50.00 ± 6.25        |
| Suero           | 1389.6 - 1528.6               | 59.27                | 44.88 ± 1.83        |

Una forma popular de analizar la calidad de modelos cuantitativos de **glucosa** es mediante la rejilla de error de Clarke [155], que se subdivide en cinco zonas (A-E). Los valores que se encuentran en las zonas A y B representan resultados precisos o aceptables; la zona C podría conducir a un mal resultado; la zona D representa una falla peligrosa para detectar y/o tratar y la zona E significa “tratamiento erróneo”. Considerando los mejores resultados obtenidos de la cuantificación de **glucosa**, la Tabla 5.20 muestra los porcentajes por zonas al usar la rejilla de error de Clarke. La figura 5.6 muestra a dicha rejilla con los resultados obtenidos en suero, para ver las rejillas de antebrazo, muñeca y dedo véase apéndice A.7, A.8, A.9.

**Tabla 5.20:** Porcentaje de aciertos por zona en la rejilla de error de Clarke.

| Espectros analizados | A       | B       | C      | D      | E   |
|----------------------|---------|---------|--------|--------|-----|
| Antebrazo            | 82.61 % | 15.22 % | 0 %    | 2.17 % | 0 % |
| Muñeca               | 47.83 % | 47.83 % | 2.17 % | 2.17 % | 0 % |
| Dedo                 | 71.74 % | 26.09 % | 0.72 % | 1.45 % | 0 % |
| Suero                | 94.78 % | 4.35 %  | 0 %    | 0.87 % | 0 % |

Comparando los resultados obtenidos con lo reportado en el estado del arte, es posible decir que a pesar que se han realizado varias investigaciones con el fin de lograr una cuantificación confiable de HbA<sub>1c</sub>, no se ha informado de una combinación como la presentada en este trabajo, la cual consiste de



**Figura 5.6:** Rejilla de error de Clarke para la predicción de glucosa en suero.

diferentes concentraciones en un población de múltiples voluntarios con 36 concentraciones diferentes; el error obtenido en este trabajo de la medición realizada en la muñeca de 46 voluntarios en un rango de 5.2 - 14 % de HbA1c fue de  $0.69 \pm 0.07$  % superando a lo reportado hasta el momento en la literatura.

Por otro lado, en el caso de glucosa *in vivo* e *in vitro* las investigaciones son abundantes como se mencionó en el estado del arte. En este trabajo se obtuvieron los siguientes resultados: en el caso de las mediciones en suero (*in vitro*) se obtuvo un error  $11.86 \pm 5.45$  mg/dL en un rango de 56 - 400 mg/dL, el cual es un error menor a los reportado por [56, 59, 60, 61, 62]. Respecto a las mediciones *in vivo* se obtuvo un error de  $30.12 \pm 0.53$  mg/dL en mediciones del antebrazo de 46 voluntarios en un rango de 56 - 400 mg/dL, con un porcentaje de acierto en la zona A de 82.61 %, 15.22 % en la zona B, para la zona D 2.17 % y 0 % para zona C y E. Comparando los resultados obtenidos con lo reportado en el estado del arte tenemos que, [66] obtuvieron un 78.4 % de aciertos en la zona A, [69] obtuvieron 72 % en la zona A, [67] obtuvieron 93 % zona A y B, sin embargo, ellos presentan un estudio por voluntario. [70] muestran una rejilla de error de Clarke sin incidencias en D, sin embargo, no expresan sus porcentajes, y su estudio fue individualizado. Por lo tanto, es posible decir que nuestro estudio aporta el incremento de aciertos en zona A, con una población de múltiples sujetos, lo cual representan resultados precisos o aceptables de medición alternativa de glucosa *in vivo*.

Sin embargo, acorde la norma ISO 15197, la Asociación América de diabetes

(ADA) y la Organización Internacional de Normalización y la Administración de Drogas y Alimentos de los EE. UU, se necesita que el 95 % de todos los valores medidos estén dentro del 20 % de exactitud de los valores medidos con un método ya conocido [156, 157], es decir, se requiere un 95 % de aciertos en la zona A de la rejilla de error. Por lo tanto, para poder aplicar esta metodología para mediciones de glucosa *in vivo*, es necesario pulirla con la finalidad de alcanzar el 95 %.

## 5.4. Clasificación

La clasificación solo se realizó en base a los porcentajes de HbA<sub>1c</sub>; puesto que se pueden utilizar para separar por clases a los voluntarios. Como se mencionó anteriormente, esta prueba se realizó con la finalidad de tener una representación más general del comportamiento de nuestra base de datos. Para ello los datos se dividieron en tres conjuntos; entrenamiento, validación y prueba, para seleccionar el porcentaje de cada conjunto se realizaron pruebas usando los conjuntos de datos de antebrazo muñeca y dedo en la región espectral de 200 a 1800 cm<sup>-1</sup>, los resultados se muestran en la tabla 5.21.

**Tabla 5.21:** Selección del porcentaje de datos para cada conjunto.

| Región corporal | % Precisión ± STD |                    |                |               |
|-----------------|-------------------|--------------------|----------------|---------------|
|                 | 60 %-20 %-20 %    | 70 %-15 %-15 %     | 80 %-10 %-10 % | 90 %-5 %-5 %  |
| Antebrazo       | 92.7 ± 2.120      | 94.05 ± 2.32       | 93.66 ± 3.64   | 93.06 ± 8.97  |
| Muñeca          | 93.45 ± 2.67      | 95.79 ± 1.06       | 94.38 ± 4.783  | 93.64 ± 5.45  |
| Dedo            | 82.68 ± 8.67      | 85.57 ± 8.20       | 82.15 ± 9.27   | 83.26 ± 12.57 |
| PROMEDIO        | 89.61 ± 6.01      | <b>91.8 ± 5.47</b> | 90.1 ± 6.86    | 90.0 ± 5.83   |

Teniendo en cuenta los resultados obtenidos de esta prueba se eligió el porcentaje de 70 % entrenamiento, 15 % validación y 15 % prueba, en la tabla 5.22 se muestran los resultados obtenidos de la clasificación usando los dos rangos espectrales (200 - 1800 cm<sup>-1</sup> y 600 - 1600 cm<sup>-1</sup>, en donde se puede observar que para ambos rangos espectrales, la región corporal que obtuvo mejores resultados es la muñeca; esto concuerda con los resultados obtenidos en regresión, y de estas dos regiones la que presenta mejor resultados es la región de 600 a 1600 cm<sup>-1</sup>.

**Tabla 5.22:** Resultados en porcentaje (%) de precisión para HbA<sub>1c</sub> en clasificación.

| Rango espectral 200 - 1800 cm <sup>-1</sup> (788 características) |              |
|---|--------------|
| Región corporal   | % precisión  |
| Antebrazo   | 91.98 ± 5.19 |
| Muñeca  | 94.75 ± 2.52 |
| Dedo  | 80.36 ± 9.18 |
| Rango espectral 600 - 1600 cm <sup>-1</sup> (512 características) |              |
| Antebrazo   | 92.58 ± 4.15 |
| Muñeca  | 96.01 ± 2.17 |
| Dedo  | 81.59 ± 9.79 |

Los resultados utilizados en la red SOM, se implementaron en el clasificador, obteniéndose lo que se muestra en la tabla 5.23. Como se observa esta representación de los datos no favoreció en el caso de clasificación; y considerando que se probaron 3 diferentes bases de datos, es posible concluir preliminarmente que la red SOM no favorece la representación de los datos de entrada, para sistemas de clasificación.

**Tabla 5.23:** Resultados en porcentaje (%) de precisión para HbA<sub>1c</sub> en clasificación usando red SOM.

| Rango espectral 200 - 1800 cm <sup>-1</sup> (788 características) |              |
|---|--------------|
| Región corporal   | % precisión  |
| Antebrazo   | 84.98 ± 1.54 |
| Muñeca  | 83.14 ± 2.90 |
| Dedo  | 80.29 ± 9.76 |
| Rango espectral 600 - 1600 cm <sup>-1</sup> (512 características) |              |
| Antebrazo   | 83.53 ± 3.31 |
| Muñeca  | 85.60 ± 1.80 |
| Dedo  | 73.04 ± 5.99 |

Seleccionando los mejores casos para cada región corporal se calculó la especificidad y sensibilidad, los resultados se muestran en la tabla 5.24, en donde se puede observar que en el caso de muñeca, tanto la sensibilidad y especificidad para las tres clases es mayor del 90 %.

Comparando los resultados obtenidos con lo ya reportado en la literatura,



**Tabla 5.24:** Especificidad y sensibilidad mejor de los casos.

| Región corporal / clase |                      | Sensibilidad ( % ) | Especificidad ( % ) |
|-------------------------|----------------------|--------------------|---------------------|
| Antebrazo               | No diabético         | 83.33              | 99.18               |
|                         | Pre-Diabético        | 87.78              | 97.51               |
|                         | Diabético            | 98.26              | 90.83               |
| Muñeca                  | <b>No diabético</b>  | <b>94.44</b>       | <b>99.73</b>        |
|                         | <b>Pre-Diabético</b> | <b>90.00</b>       | <b>98.45</b>        |
|                         | <b>Diabético</b>     | <b>98.61</b>       | <b>92.74</b>        |
| Dedo                    | No diabético         | 25.00              | 98.26               |
|                         | Pre-Diabético        | 70.00              | 95.96               |
|                         | Diabético            | 95.83              | 60.00               |

se tiene que: la referencia [50], realizó el cálculo de especificidad y sensibilidad de manera binaria; su mejor resultado se obtiene en voluntarios sanos contra diabéticos controlados con un 100 % de especificidad y sensibilidad. En nuestro caso, se calcularon estos parámetros en clasificación multiclase, donde se obtuvieron los mejores resultados en la clase no diabéticos, con un 94.44 % sensibilidad y 99.73 % especificidad. Por otro lado, de la referencia [103] distinguen entre voluntarios sanos y diabéticos usando ANN y su mejor resultado fue de 96 % de precisión y sus valores de sensibilidad y especificidad fueron 88.9 % a 90.9 %, respectivamente; en nuestro caso de manera similar se obtuvo un 96.01 % de precisión, a pesar de que la diferencia entre porcentajes de precisión no es significativa ( $P > 0.05$ ,  $P = 0.31$  usando el método no paramétrico Kruskal Wallis), se observa una mejora en los porcentajes de sensibilidad y especificidad, aunque se debe considerar que lo reportado por [103] es obtenido mediante una clasificación binaria, mientras que lo reportado en este trabajo es una clasificación multiclase.

A continuación, se presentan tablas comparativas con los resultados reportados hasta la fecha en el estado del arte, en función de las diferentes sustancias analizadas y lo obtenido por el método propuesto en este trabajo. La tabla 5.25 presenta la comparación con el estado del arte de los trabajos que determinan HbA<sub>1c</sub> *in vivo*. La tabla 5.26 muestra la comparativa con lo reportado en el estado del arte referente a mediciones de glucosa de manera *in vivo*.

La tabla 5.27 muestra la comparativa con lo reportado en el estado del arte

**Tabla 5.25:** Comparación de métodos usados para mediciones *in vitro* de HbA<sub>1c</sub>.

| Autor y año                              | Resultados  | Métodos usados   | Rango de valores  |
|--|---|------------------|---|
| J. F. Villa-Manríquez et al. (2017) [50] | Sensibilidad entre 100 y 100<br>Especificidad entre 100 y 40            | PCA-SVM          | Sano<br>Diabético controlado<br>Diabético no controlado |
| E. Guevara et al. (2018) [103]           | 96 % de clasificación<br>Sensibilidad y especificidad<br>88.9 y 90.09 % | PCA-ANN          | Diabético<br>No diabético                               |
| GIOB (2021)                              | 0.69 % RMSECV   | RReliefF-SOM-ANN | 5.2 – 14 %  |

**Tabla 5.26:** Comparación de métodos usados para mediciones *in vivo* de glucosa.

| Autor y año                           | Métodos usados   | Error             | CEG-A        | Rango de valores (mg/dL) |
|---------------------------------------|------------------|-------------------|--------------|--------------------------|
| J. Chaiken et al. (2005) [69]         | PLS              | No reportado      | 72 %         | 50 – 300                 |
| A. M. K. Enejder et al. (2005) [70]   | PLS              | 7.8 % MAE         | No reportado | 68 – 223                 |
| Scholtes-Timmerman et al. (2014) [81] | PLS              | 21.62 mg/dL MAE   | 78.4 %       | 64 – 398                 |
| Lundsgaard-Nielsen et al. (2018) [67] | MLR              | No reportado      | 54.1 %       | 72 – 540                 |
| GIOB (2021)                           | RReliefF-SOM-ANN | 30.1 mg/dL RMSECV | 82.61 %      | 56 – 400                 |

referente a las mediciones de glucosa en suero humano ( mediciones *in vitro*).

**Tabla 5.27:** Comparación de métodos usados para mediciones *in vitro* de glucosa.

| Autor y año                              | Métodos usados   | Error (mg/dL)               | Rango de valores (mg/dL) |
|--|------------------|-----------------------------|--------------------------|
| A. J. Berger et al. (1999) [61]          | PLS              | RMSEP = 26                  | 50 – 450                 |
| J. Y. Qu et al. (1999) [62]              | PLS              | RMSEP= 23.4                 | 72 – 360                 |
| D. Rohleder et al. (2004, 2005) [59, 60] | PLS              | RMSEP= 17.1<br>RMSECV= 22.6 | 42 – 423                 |
| L. Silveira et al. (2017) [56]           | PLS              | RMSECV= 26.4                | 80–120                   |
| Giansante, S., (2021) [64]               | MLR              | RMSECV= 15.4                | 70-258                   |
| GIOB (2021)                              | RReliefF-SOM-ANN | RMSECV= 11.9                | 56 – 400                 |

## 5.5. Análisis espectral

A continuación, se presenta el análisis espectral realizado a muestras puras de glucosa, HbA<sub>1c</sub> y a las mediciones *in vivo* e *in vitro* realizadas a los voluntarios. En el caso de HbA<sub>1c</sub> se identificaron 13 picos representativos que se pudieron corroborar con lo obtenido en la literatura [1], los cuales se muestran en la tabla 5.28.

Los picos encontrados en la sustancia pura, se buscaron en los espectros *in vivo* de los voluntarios. Para considerar que los picos pertenecen a la sustancia en análisis, se consideró un error de  $\pm 10 \text{ cm}^{-1}$  del valor reportado en la

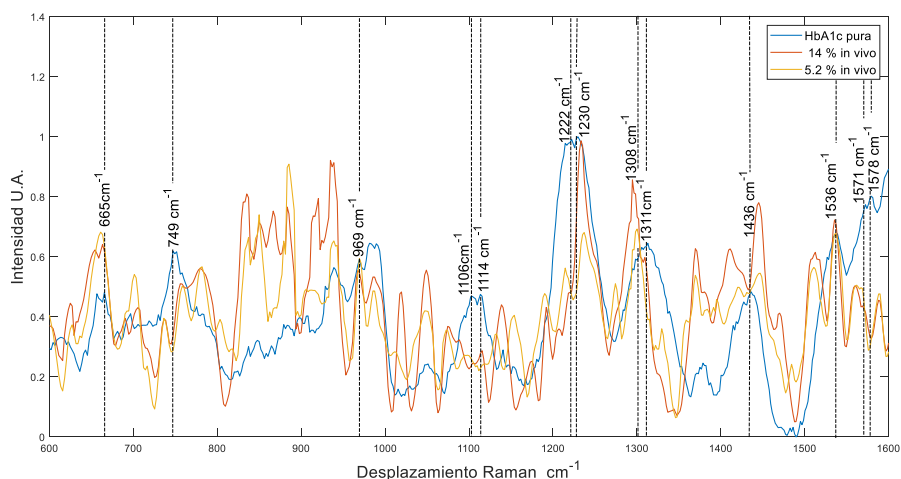
**Tabla 5.28:** Asignación molecular de picos de HbA<sub>1c</sub> en polvo, basada en otros trabajos.

| Posiciones experimental (cm <sup>-1</sup> ) | Posiciones reportadas (cm <sup>-1</sup> )             | Asignación molecular  |
|---|---|---|
| 1578  | 1585 [158], 1582 [159, 160], 1580 [161]               | $\nu_{37} \nu(C_{\alpha}C_m)_{sym}$ [158, 159, 161], Fenilalanina [161], $\nu_{37} \nu(C_{\alpha}C_m)_{asym}$ [160]   |
| 1571  | 1568 [158, 162]                                       | $\nu(C_{\alpha}C_m)_{asym}$ [158, 162]  |
| 1536  | 1548 [158], 1539-1546 [154, 163], 1544 [160]          | Amida II ( $\beta$ -giro) dentro y fuera de fase combinación de C-N estiramiento y flexión en plano NH, [154, 163] $\nu(C_{\beta}C_{\beta})$ [158] $\nu_{11} \nu(C_{\beta}C_{\beta})$ [160]         |
| 1436  | 1427 [158], 1430 [160]                                | $\nu_{28} \nu(C_{\alpha}C_m)_{sym}$ [158]<br>$\nu_{28} \nu(C_{\alpha}C_m)_{sym}$ [160]  |
| 1311  | 1310 [164]  | torsión [164]   |
| 1308  | 1303 [158], 1306 [160]                                | $\delta_{asym}(C_mH)$ [158], $\nu_{21} \delta_{asym}(C_mH)$ [160]   |
| 1230  | 1228 [161], 1231 [162], 1230 [160]                    | $\delta(C_mH)$ [161], $\nu_{13}$ [162], metRBC $\delta(C_mH)$ [160]   |
| 1222  | 1226 [158], 1222 [159], 1223-1225 [163], 1225 [160]   | $\delta(C_mH)$ [158, 159, 160], estado oxigenado de la hemoglobina, PO <sub>2</sub> vibraciones de estiramiento asimétrico, ácidos nucleicos [163]  |
| 1114  | 1120 [158], 1122-1128 [163]                           | $\nu_{22}$ , $\delta(CH_2)$ , torsión, agitación [158], Estiramiento C-N en proteínas, glucosa, C-metilo en hemo [163]  |
| 1106  | 1106 [154]  | glucosa C-O-C (deformación del modelo de flexión) [154]   |
| 969   | 974 [159], 970 [161], 976 [163, 154], 975 [160]       | $\gamma(C_4H=)$ [159], p: vibración esquelética, glutación [161], CH <sub>3</sub> deformación [163], C-O ángel de flexión de glucosa [154], $\nu_{46} \delta(pyr \text{ deformación})_{asym}$ [160] |
| 749   | 755 [158], 754 [159], 753 [161, 160], 754 [163]       | $\nu(pyr \text{ deformación})$ [159], Estiramiento pyr. respiración de oxihemoglobina [163, 161], $\nu(pyr \text{ respiración})$ $\nu_{15}$ [160]   |
| 665   | 677 [158], 676 [161], 673 [159], 672 [163], 664 [160] | $\nu_7 \delta(pyr \text{ deformación})_{sym}$ [158, 161, 159, 160] estado desoxigenado de la hemoglobina [163]  |

sustancia pura y que en los 36 valores de HbA<sub>1c</sub> se encontrara presente ese pico. Las tablas que se obtuvieron se muestran en los apéndices A.2, A.1, A.3 ; en ellas se observa que los picos que se mantienen constantes en las muestras *in vivo* son: para el antebrazo, 1536, 1230, 1114, 969 y 665 cm<sup>-1</sup> para la muñeca 1536, 1230 y 665 cm<sup>-1</sup> y para el dedo 1536 y 1308 cm<sup>-1</sup>. Es interesante notar que, a pesar de tener un mayor número de picos localizados en la región del antebrazo, no fue la región corporal que presentó el mejor resultado, por otro lado, el dedo presentó solo dos picos de la sustancia pura y entre ellos no

aparece el pico más intenso de la sustancia pura.

En la figura 5.7 se puede observar una representación visual de tres espectros Raman y sus respectivos ajustes gaussianos, estos espectros corresponden a la sustancia pura y dos concentraciones de HbA1c *in vivo*. En el caso de la glucosa pura se encontraron los picos reportados en la tabla



**Figura 5.7:** Picos Raman de HbA1c pura comparados con 14 % y 5.2 % *in vivo*.

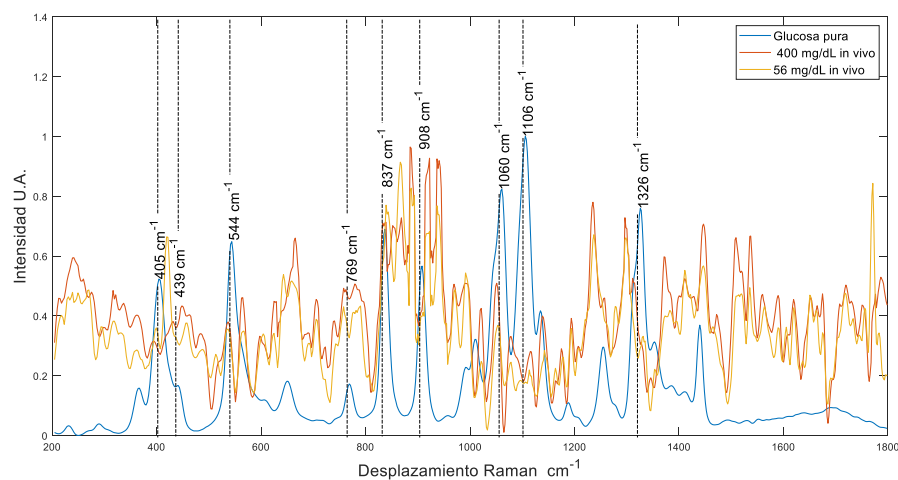
5.29, de igual manera que para la HbA1c se presentan aquellos picos que fue posible corroborar con la literatura.

**Tabla 5.29:** Asignación molecular de picos de glucosa en polvo basada en otros trabajos.

| Posiciones experimental (cm <sup>-1</sup> ) | Posiciones reportadas (cm <sup>-1</sup> )  | Asignación molecular   |
|---|--|--|
| 1353  | 1346 [165], 1365 [166],                    | deformaciones $\delta$ CH <sub>2</sub> y $\delta$ CH <sub>2</sub> OH [165]   |
| 1106  | 1120 [165], 1106 [154], 1106 [167]         | C-O-C (flexión modelo de deformación) [154, 104]   |
| 1060  | 1060 [65], 1054 [165], 1065 [166]          | Estiramiento $\nu$ (C-O) y $\nu$ (C-C) [165, 104]  |
| 908   | 914 [165], 912 [166]                       | $\delta$ (COH), $\delta$ (CCH) y $\delta$ (OCH) [165]  |
| 837   | 841 [165], 854 [166], 838 [104], 838 [167] | $\delta$ (COH), $\delta$ (CCH) y $\delta$ (OCH) [165], vibraciones $\nu$ (C-C) [167]   |
| 769   | 759 [65], 772 [165], 776 [104]             | $\nu$ (C-C) y $\delta$ (C <sub>1</sub> -H <sub>1</sub> ) [104]   |
| 544   | 542 [165]                                  | Deformación exocíclica [165]   |
| 439   | 440 [165], 436 [166], 437 [104]            | $\delta$ (C <sub>2</sub> -C <sub>1</sub> -O <sub>1</sub> ) flexión de la vibración en $\alpha$ -glucosa y $\beta$ -glucosa [104] |
| 405   | 405 [165]                                  | Deformaciones endocíclicas [165]   |

De los espectros *in vivo*, se tienen 43 concentraciones diferentes. Después de realizar la identificación de picos a cada una de estas concentraciones se

encontró que para el antebrazo, los picos que se mantienen constante en todas las concentraciones fueron 544, 837 y 1060  $\text{cm}^{-1}$ , en el caso de la muñeca 544 y 837  $\text{cm}^{-1}$ , y para el dedo 544 y 837  $\text{cm}^{-1}$ , en el caso del suero 544, 837 y 1060  $\text{cm}^{-1}$ . En la figura 5.8 se observa el espectro de la glucosa pura, así como el espectro de 2 mediciones *in vivo* de la muñeca, para ejemplificar como se localizaron los picos en las mediciones en voluntarios.



**Figura 5.8:** Picos Raman de glucosa pura comparados con 56 mg/dL y 400 mg/dL *in vivo*.

Otro análisis que se realizó fue el cálculo de la relación señal a ruido (SNR) de cada uno de los 46 voluntarios. La SNR en los espectros Raman se calculó usando la ecuación 5.1, la cual describe la relación de la media del pico más intenso  $\bar{S}$  y la desviación estándar a dicha frecuencia  $\sigma_y$  [92]. El cálculo se realizó utilizando el pico a  $1230\text{ cm}^{-1}$  para HbA1c y  $1106\text{ cm}^{-1}$  para glucosa, usando nueve mediciones por individuo.

$$SNR = \frac{\bar{S}}{\sigma_y} \quad (5.1)$$

Como se observa en la tabla 5.30, la SNR de la glucosa es menor que la de la HbA1c, sin embargo, se obtuvo una  $SNR \geq 3$ , por lo tanto es posible realizar los análisis cuantitativos [92] que se presentaron a lo largo de este trabajo. En el caso del suero solo se calculó en función de glucosa, dado que esa fue la sustancia de importancia para el análisis.

**Tabla 5.30:** Relación señal a ruido (SNR) promedio de los 46 voluntarios.

| Región analizada | SNR promedio (46 voluntarios)   |                                   |
|------------------|---------------------------------|-----------------------------------|
|                  | HbA1c ( $1230\text{ cm}^{-1}$ ) | glucosa ( $1106\text{ cm}^{-1}$ ) |
| Antebrazo        | $9.75 \pm 7.26$                 | $4.52 \pm 1.70$                   |
| Muñeca           | $11.63 \pm 9.65$                | $5.02 \pm 2.34$                   |
| Dedo             | $17.06 \pm 29.77$               | $8.28 \pm 13.59$                  |
| Suero            | NA                              | $12.43 \pm 5.23$                  |

## 5.6. Resumen de capítulo

En este capítulo se presentaron los resultados de la cuantificación no invasiva de HbA1c y glucosa mediante la combinación de espectroscopia Raman, redes neuronales artificiales y métodos de selección y extracción de características, así como de mejoramiento espectral mediante la red SOM, mostrando con ello la factibilidad de utilizar espectroscopia Raman para dichas mediciones. En el siguiente capítulo se presentarán las conclusiones de este trabajo de manera más detallada.

---

## CONCLUSIONES Y TRABAJO FUTURO

---

En este trabajo se realizó la caracterización de HbA<sub>1c</sub> pura y de glucosa, usando espectroscopia Raman, así mismo la asignación molecular de picos acorde con lo reportado a la literatura. En particular, se realizaron diferentes concentraciones de las sustancias y se utilizaron métodos de regresión y clasificación basados en ANN con la finalidad de determinar y aproximar dichas concentraciones.

Cabe resaltar que la potencia del láser se eligió en función de la sustancia analizada, podría decirse que mediante experimentación. Por ejemplo, la hemoglobina glicosilada pura, con más de 10mW se quemaba, el detector se saturaba y ya no mostraba picos, esto debido a su color, la hemoglobina pura es roja y absorbe. En el caso del suero al ser un líquido amarillo semi transparente fue posible incrementar la potencia sin dañar la muestra, y generalmente mientras mayor sea la potencia (sin llegar a dañar la muestra) se obtiene una señal menos ruidosa y con picos más definidos. En el caso de las mediciones *in vivo*, se consideró la norma ANSI para uso de láseres en piel y se realizó el cálculo de la potencia y tiempo de exposición máximos para evitar dañar a los voluntarios, y se obtuvo que 60mW estaba dentro del cálculo de la normal y la señal de los espectros era mejor que con menores potencias, aunque aun así la señal es sumamente ruidosa. Si se quisieran realizar mediciones con 10mW de potencia (potencia usada para mediciones de la hemoglobina pura) *in vivo* se obtendría una señal ruidosa y sería difícil o casi imposible determinar algún pico de las sustancias de interés. En conclusión, la potencia juega un papel muy importante en los resultados, puesto que si no se tiene una señal Raman que represente a la sustancia de interés, no se obtendrá una predicción de valores adecuada, por lo tanto, la diferencia entre valor real y predicho sería mayor al

obtenido en este trabajo.

Para la preparación de los datos se probaron diferentes métodos de extracción y selección de características, así como acondicionamiento de la señal para mejorar la aproximación de concentraciones de glucosa en suero de 50 voluntarios diabéticos proporcionados por el IMSS-CIBIOR. Estos sueros se midieron utilizando un contenedor de polipropileno, se utilizó la red SOM y ANN, con una división de datos de 90 % de entrenamiento y 10 % prueba se obtuvo un error de 13.50 mg/dL. Por otro lado, se midieron sueros de 46 voluntarios diabéticos y no diabéticos; en este caso se utilizó un contenedor de aluminio con la finalidad de disminuir el error producido por las contribuciones en bandas Raman que genera el polipropileno, en este caso se realizó una validación cruzada de 4 particiones para evitar el sobre ajuste y se obtuvo un error y desviación estándar de  $11.86 \pm 5.45$  mg/dL utilizando la red SOM y RReliefF para seleccionar características (150 características seleccionadas) y ANN para realizar la aproximación, con esto se puede observar que se obtuvo una mejora a lo reportado en el estado del arte.

Con respecto a las mediciones *in vivo*, se obtuvo un error y desviación estándar de  $0.69 \pm 0.07$  para determinar la concentración de HbA<sub>1c</sub> en 46 voluntarios diabéticos y no diabéticos, en un rango de 5.2 - 14 % de HbA<sub>1c</sub>. Este resultado se logró al analizar los espectros obtenidos de la muñeca de los voluntarios mediante la implementación de RReliefF como selector de características (512 características seleccionadas), red SOM y ANN para realizar la aproximación. Como se observa en el estado de arte esto no se había implementado anteriormente con una población no controlada, es decir, en donde se desconoce el historial clínico de los voluntarios y con un experimento intra-sujetos.

Por otro lado, en el caso de glucosa *in vivo* se obtuvo un error y desviación estándar de  $30.12 \pm 0.53$  mg/dL, en mediciones del antebrazo de 46 voluntarios, en un rango de 56 - 400 mg/dL; con un porcentaje de acierto en la rejilla de error de Clarke de 82.61 % en la zona A, 15.22 % en la zona B, para la zona D 2.17 % y 0 % para zona C y E. Este resultado se obtuvo utilizando RReliefF (200 características seleccionadas), red SOM y ANN para la regresión. Los resultados presentan una disminución del error reportado en el estado del arte considerando el tipo de población analizada.



En resumen el mejor lugar para medición en el caso de la HbA<sub>1c</sub> fue en la muñeca y para el caso de la glucosa fue el antebrazo, en estos dos lugares se obtuvo la menor diferencia entre valores reales y valores predichos por la red neuronal artificial, el hecho de que para cada sustancia el lugar de medición sea diferente, puede deberse a que la selección características que representan a cada sustancia varió en función de la salida esperada, es decir, estas características se seleccionaron considerando la influencia de los valores de laboratorio, esto con la finalidad de tener una mejor representación del problema, ya que el algoritmo tiene que aprender de manera adecuada para poder predecir correctamente en el futuro. Cabe destacar que obtener la mejor representación de los espectros Raman a partir de la selección y/o extracción de características fue un reto, debido a que la SNR de algunas mediciones fue  $\leq 3$ , complicando con ello la cuantificación de las sustancias analizadas *in vivo*.

A pesar de los prometedores resultados obtenidos, esta metodología aun no puede ser implementada en hospitales, debido a que es necesario incrementar la población de estudio con la finalidad de lograr una adecuada generalización de la técnica. Sin embargo, es posible concluir que mediante el uso de espectroscopia Raman, en conjunto con técnicas de selección de características y el uso de redes neuronales artificiales es posible la cuantificación preliminar de HbA<sub>1c</sub> y glucosa *in vivo*, como una medición alternativa y no invasiva.

## 6.1. Contribuciones

Las principales contribuciones de esta investigación son:

- Una caracterización de HbA<sub>1c</sub> pura usando espectroscopia Raman y modelos de regresión basados en ANN.
- Un proceso de cuantificación innovador basado en redes neuronales artificiales para cuantificar HbA<sub>1c</sub> *in vivo* y glucosa en suero humano.
- Un análisis sistemático de algoritmos de selección de características obtenidas de espectros Raman.
- El uso exitoso de mapas auto-organizados (SOM por sus siglas en inglés) neuronales para el refinamiento de características obtenidas de señales

espectrales Raman.

## 6.2. Trabajo futuro

Como trabajo futuro se propone:

- Implementar los algoritmos presentados en este trabajo en una población controlada, es decir, en la cual se conozca su historial clínico, así como aplicar dicha metodología por sujeto, analizando sus espectros a lo largo del tiempo, esto con la finalidad de descartar errores generados por mediciones en múltiples sujetos.
- Analizar por sexo, edad y peso de la actual base de datos.
- Modificar los hiper-parámetros de las redes neuronales, tales como funciones de activación, algoritmos de entrenamiento, entre otros, con la final de disminuir el error de predicción.
- Utilizar otros tipos de selectores de características como arboles de decisión u otros métodos embebidos.
- Incrementar el tamaño del spot y la potencia del láser sin dañar el tejido para obtener una mejora en la SNR y con ello disminuir el error en la predicción.

### 6.3. Trabajos y participaciones derivadas de esta tesis

- N. **González-Viveros**, J. Castro-Ramos, P. Gómez-Gil, H.H. Cerecedo-Núñez, "On the estimation of sugars concentrations sing Raman spectroscopy and artificial neural networks". Food Chemistry, page 119077, 2021. ISSN 1386-1425. [https : //doi.org/10.1016/j.foodchem.2021.129375](https://doi.org/10.1016/j.foodchem.2021.129375). **Artículo**.
- N. **González-Viveros**, J. Castro-Ramos, P. Gómez-Gil, H.H. Cerecedo-Núñez, "Characterization of glycated hemoglobin based on Raman spectroscopy and artificial neural networks", Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, Volume 247, 2021, 119077, ISSN 1386-1425, [https : //doi.org/10.1016/j.saa.2020.119077](https://doi.org/10.1016/j.saa.2020.119077). **Artículo**.
- N. Vázquez Osorio, F. Gutiérrez-Delgado, N. **González-Viveros**, C. Villalba-Hernández,, F. Narea- Jiménez, JL Flores-Guerrero, and J. Castro-Ramos, "Integral Analysis To Detect Type 2 Diabetes Using Biomarkers and Raman Spectroscopy", poster presentation in OSA Frontiers in Optics + Laser Science, Septembter 2020. **Memoria de congreso**.
- Flores-Guerrero, J.L., Muñoz-Morales, A., Narea-Jimenez, F.; Perez-Fuentes, R., Torres-Rasgado, E., Ruiz-Vivanco, G., **Gonzalez-Viveros**, N., Castro-Ramos, J., "Novel Assessment of Urinary Albumin Excretion in Type 2 Diabetes Patients by Raman Spectroscopy", Diagnostics 2020, 10, 141, [https : //doi.org/10.3390/diagnostics10030141](https://doi.org/10.3390/diagnostics10030141). **Artículo**.
- N. López-Hernández, N. **González-Viveros**, H.N. Chavarría-Lizárraga, J.J. Sánchez-Escobar, J. Castro-Ramos "Espectros Raman de diferentes concentraciones acuosas de glucosa, sacarosa y fructosa", LXII Congreso Nacional de Física, Octubre 2019. **Poster**.
- Castro-Ramos J., Narea-Jiménez F., Muñoz-Morales A., Gómez-Gil, P.,

**González-Viveros, N.**, Sánchez-Escobar J.J., Muñoz-López, J., “Detecting concentrations of Glucose, Sucrose and D-fructose in solutions by Means of Raman Spectroscopy, Support Vector Machine and Principal Component Analysis”; FACSS and Scix conference organization 2019, Palm Springs, California. **Poster.**

- **Castro-Ramos J.**, Narea-Jiménez F., Muñoz-Morales A., Gómez-Gil, P., **González-Viveros, N.**, Sánchez-Escobar J.J., Muñoz-López, J., “Detecting glucose, sucrose and fructose levels by means of Raman spectroscopy, support vector machine and principal component analysis”, FACSS and Scix conference organization 2018, Atlanta, Georgia. **Poster.**
- **Castro-Ramos, J.**, Villa-Manríquez, J.F., Gómez-Gil, P., **González-Viveros, N.**, Narea-Jiménez, F., Sánchez-Escobar, J.J., Muñoz-López J., “Detecting glucose levels by means of Raman spectroscopy, support vector machine and principal component analysis”. ECORFAN Journal Taiwan. 2017 1(2): 11–15. **Capítulo de libro.**

### Trabajos pendientes

- **N. González-Viveros**, J. Castro-Ramos, P. Gómez-Gil, H.H. Cerecedo-Núñez, R. Perez-Fuentes, E. Torres-Rasgado, F. Gutierrez-Delgado. “Estimation of glucose concentration in human serum using Raman Spectroscopy and Artificial Neural Networks”. **Artículo en proceso de redacción.**
- **N. González-Viveros**, J. Castro-Ramos, P. Gómez-Gil, H.H. Cerecedo-Núñez, F. Gutierrez-Delgado, E. Torres-Rasgado, R. Pérez-Fuentes, J. L Flores-Guerrero. ‘Quantification of glycated hemoglobin and glucose in vivo using Raman spectroscopy and Artificial Neural Networks: a pilot study.’. **Artículo en proceso de redacción.**

---

## BIBLIOGRAFÍA

---

- [1] N. González-Viveros, J. Castro-Ramos, P. Gómez-Gil, and H. H. Cerecedo-Núñez. Characterization of glycated hemoglobin based on raman spectroscopy and artificial neural networks. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, page 119077, 2020. ISSN 1386-1425. doi:[10.1016/j.saa.2020.119077](https://doi.org/10.1016/j.saa.2020.119077).
- [2] M. Bracho-Nava, V Stepenska-Alvarez, M Sindas-villasmil, Yoleida rivas de casal, María Bozo de GoNzález, and anyelo duran Mojica. Hemoglobina glicosilada o hemoglobina glicada, ¿cuál de las dos? *SABER. Revista Multidisciplinaria del Consejo de Investigación de la Universidad de Oriente*, 27:521–529, 2015. ISSN 1315-0162. URL <https://www.redalyc.org/articulo.oa?id=427744808002>.
- [3] G. R. Desiraju. The Raman Effect. *Nature India*, pages 1–3, 2008. ISSN 1755-3180. doi:[10.1038/nindia.2008.302](https://doi.org/10.1038/nindia.2008.302).
- [4] B&W tek. Theory of raman scattering, 2013. URL <https://bwtek.com/raman-theory-of-raman-scattering/>. visited on 2020-09-10.
- [5] D. Frackowiak. The jablonski diagram. *Journal of Photochemistry and Photobiology B: Biology*, 2(3):399, 1988. ISSN 1011-1344. doi:[10.1016/1011-1344\(88\)85060-7](https://doi.org/10.1016/1011-1344(88)85060-7).
- [6] D. W. Shipp, F. Sinjab, and I. Notingher. Raman spectroscopy: techniques and applications in the life sciences. *Adv. Opt. Photon.*, 9(2):315–428, Jun 2017. doi:[10.1364/AOP.9.000315](https://doi.org/10.1364/AOP.9.000315).
- [7] S. Haykin. *Neural Networks and Learning Machines*. Prentice Hall, 3 edition, 2009. ISBN 9780131471399.

- 
- [8] MathWorks. Hyperbolic tangent sigmoid transfer function, 2020. URL <https://www.mathworks.com/help/deeplearning/ref/tansig.html>. Accessed 21 October 2020.
- [9] MathWorks. Linear transfer function, 2020. URL <https://www.mathworks.com/help/deeplearning/ref/purelin.html>. Accessed 21 October 2020.
- [10] T. Alaa. Classification assessment methods. *Applied Computing and Informatics*, jan 2020. doi:[10.1016/j.aci.2018.08.003](https://doi.org/10.1016/j.aci.2018.08.003).
- [11] Eklavya. Kohonen self-organizing maps, 2019. URL <https://towardsdatascience.com/kohonen-self-organizing-maps-a29040d688da>. Accessed 21 October 2020.
- [12] M. A. Hall. *Correlation-based Feature Subset Selection for Machine Learning*. PhD thesis, University of Waikato, Hamilton, New Zealand, 1998.
- [13] William L Clarke, Daniel Cox, Linda A Gonder-Frederick, William Carter, and Stephen L Pohl. Evaluating clinical accuracy of systems for self-monitoring of blood glucose. *Diabetes Care*, 10(5):622–628, 1987. ISSN 0149-5992. doi:[10.2337/diacare.10.5.622](https://doi.org/10.2337/diacare.10.5.622). URL <https://care.diabetesjournals.org/content/10/5/622>.
- [14] American Diabetes Association. Diagnosis and classification of diabetes mellitus. *Diabetes Care*, 36, 2013. ISSN 0149-5992. doi:[10.2337/dc13-S067](https://doi.org/10.2337/dc13-S067).
- [15] World Health Organization. Use of glycated haemoglobin (HbA<sub>1c</sub>) in the diagnosis of diabetes mellitus: abbreviated report of a WHO consultation. *World Health Organization 2011*, pages 1–25, 2011.
- [16] L. Liu, S. Hood, Y. Wang, R. Bezverkov, C. Dou, A. Datta, and C. Yuan. Direct enzymatic assay for % HbA<sub>1c</sub> in human whole blood samples. *Clinical Biochemistry*, 41:576–583, 2008. doi:[10.1016/j.clinbiochem.2008.01.013](https://doi.org/10.1016/j.clinbiochem.2008.01.013).
- [17] H. F. Bunn, D. N. Haney, S. Kamin, K. H. Gabbay, and P. M. Gallop. The biosynthesis of human hemoglobin A<sub>1c</sub>. Slow glycosylation of

- hemoglobin in vivo . *The Journal of Clinical Investigation*, 57(6):1652–1659, 1976. doi:[10.1172/JCI108436](https://doi.org/10.1172/JCI108436).
- [18] S. I. Sherwani, H. A. Khan, A. Ekhzaimy, A. Masood, and M. K. Sakharkar. Significance of HbA1c test in diagnosis and prognosis of diabetic patients. *Biomarker insights*, 11:95–104, 2016. doi:[10.4137/BMI.S38440](https://doi.org/10.4137/BMI.S38440).
- [19] M. Abe and K. Matsumoto. Glycated hemoglobin or glycated albumin for assessment of glycemic control in hemodialysis patients with diabetes? *nature clinical practice: nephrology*, 4(9):482–483, 2008. doi:[10.1038/ncpneph0881](https://doi.org/10.1038/ncpneph0881).
- [20] H. Zheng, J. He, P. Li, M. Guo, H. Jin, J. Shen, Z. Xie, and C. Chi. Glucose Screening Measurements and Noninvasive Glucose Monitor Methods. *Procedia Computer Science*, 139:613–621, 2018. ISSN 1877-0509. doi:[10.1016/j.procs.2018.10.202](https://doi.org/10.1016/j.procs.2018.10.202).
- [21] Medlineplus, 2020. URL <https://medlineplus.gov/spanish/a1c.html>. Accessed 21 October 2020.
- [22] X. Zhang, K. F. Medzihradszky, J. Cunningham, P. D. Lee, C. L. Rognerud, C. N. Ou, P. Harmatz, and H. E. Witkowska. Characterization of glycated hemoglobin in diabetic patients : usefulness of electrospray mass spectrometry in monitoring the extent and distribution of glycation. *Journal of Chromatography B: Biomedical Sciences and Applications*, 759(1): 1–15, 2001. doi:[10.1016/S0378-4347\(01\)00196-7](https://doi.org/10.1016/S0378-4347(01)00196-7).
- [23] Encyclopedia Britannica, 2020. URL <https://www.britannica.com/science/glucose>. Accessed 21 October 2020.
- [24] R. Steane. Structure of glucose and other carbohydrate molecules, 2020. URL <https://www.biotopics.co.uk/as/glucose2.html>. Accessed 21 October 2020.
- [25] World Health Organization, 2020. URL [https://www.who.int/diabetes/publications/Definition%20and%20diagnosis%20of%20diabetes\\_.pdf](https://www.who.int/diabetes/publications/Definition%20and%20diagnosis%20of%20diabetes_.pdf).

- [26] S. H. Ang, M. Rambeli, T. M. Thevarajah, Y. B. Alias, S. M. Khor. Quantitative, single-step dual measurement of hemoglobin A<sub>1c</sub> and total hemoglobin in human whole blood using a gold sandwich immunochromatographic assay for personalized medicine. *Biosensors and Bioelectronics*, 78:187 – 193, 2016. ISSN 0956-5663. doi:[10.1016/j.bios.2015.11.045](https://doi.org/10.1016/j.bios.2015.11.045).
- [27] S. H. Ang, T. M. Thevarajah, P. M. Woi, Y. B. Alias, S. M. Khor. A lateral flow immunosensor for direct, sensitive, and highly selective detection of hemoglobin A<sub>1c</sub> in whole blood. *Journal of chromatography. B*, 1015-1016: 157—165, March 2016. ISSN 1570-0232. doi:[10.1016/j.jchromb.2016.01.059](https://doi.org/10.1016/j.jchromb.2016.01.059).
- [28] H. Sirén, P. Laitinen, U. Turpeinen, and P. Karppinen. Direct monitoring of glycohemoglobin A<sub>1c</sub> in the blood samples of diabetic patients by capillary electrophoresis: Comparison with an immunoassay method. *Journal of Chromatography A*, 979(1):201 – 207, 2002. ISSN 0021-9673. doi:[10.1016/S0021-9673\(02\)01403-6](https://doi.org/10.1016/S0021-9673(02)01403-6).
- [29] E. del Castillo, M. Montes-Bayón, E. Añón, and A. Sanz-Medel. Quantitative targeted biomarker assay for glycated haemoglobin by multidimensional LC using mass spectrometric detection. *Journal of Proteomics*, 74(1):35 – 43, 2011. ISSN 1874-3919. doi:[10.1016/j.jprot.2010.07.011](https://doi.org/10.1016/j.jprot.2010.07.011).
- [30] S.-. Lee, L.-H. Wang, S.-M. Tsai, H.-Y. Fang, and L.-Y. Tsai. Effects of the Hb E, Hb H and Hb G-Taichung variants on HbA<sub>1c</sub> values by the Bio-Rad variant II turbo analyzer. *Clinical biochemistry*, 44(16):1338—1342, November 2011. ISSN 0009-9120. doi:[10.1016/j.clinbiochem.2011.08.907](https://doi.org/10.1016/j.clinbiochem.2011.08.907).
- [31] M. Adamczyk, Y.-Y. Chen, D. D. Johnson, P. G. Mattingly, J. A. Moore, Y. Pan, and R. E Reddy. Chemiluminescent acridinium-9-carboxamide boronic acid probes: Application to a homogeneous glycated hemoglobin assay. *Bioorganic & Medicinal Chemistry Letters*, 16(5):1324—1328, March 2006. ISSN 0960-894X. doi:[10.1016/j.bmcl.2005.11.062](https://doi.org/10.1016/j.bmcl.2005.11.062).
- [32] K. Hirokawa, K. Nakamura, and N. Kajiya. Enzymes used for the



- determination of HbA<sub>1c</sub>. *FEMS Microbiology Letters*, 235(1):157–162, 2004. doi:[10.1111/j.1574-6968.2004.tb09581.x](https://doi.org/10.1111/j.1574-6968.2004.tb09581.x).
- [33] O. S. Zhernovaya, V. V. Tuchin, I. V. Meglinski. Monitoring of blood proteins glycation by refractive index and spectral measurements. *Laser Physics Letters*, 5(6):460–464, 2008. doi:[10.1002/lapl.200810007](https://doi.org/10.1002/lapl.200810007).
- [34] M. . de la Calle Guntiñas, R. Wissiack, G. Bordin, and A. Rosa Rodríguez. Determination of haemoglobin A<sub>1c</sub> by liquid chromatography using a new cation-exchange column. *Journal of Chromatography B*, 791(1):73 – 83, 2003. ISSN 1570-0232. doi:[10.1016/S1570-0232\(03\)00202-2](https://doi.org/10.1016/S1570-0232(03)00202-2).
- [35] A. Maleska, C. Hirtz, E. Casteleyn, O. Villard, J. Ducos, A. Avignon, A. Sultan, and S. Lehmann. Comparison of HbA<sub>1c</sub> detection in whole blood and dried blood spots using an automated ion-exchange HPLC system. *Bioanalysis*, 9(5):427–434, 2017. doi:[10.4155/bio-2016-0278](https://doi.org/10.4155/bio-2016-0278).
- [36] Y. Li, J.-O. Jeppsson, M. Jörntén-Karlsson, E. L. Larsson, H. Jungvid, I. Y. Galaev, and B. Mattiasson. Application of shielding boronate affinity chromatography in the study of the glycation pattern of haemoglobin. *Journal of Chromatography B*, 776(2):149 – 160, 2002. ISSN 1570-0232. doi:[10.1016/S1570-0232\(02\)00162-9](https://doi.org/10.1016/S1570-0232(02)00162-9).
- [37] M. Thevarajah, M. N. Nadzimah, and Y. Y. Chew. Interference of hemoglobinA<sub>1c</sub> (HbA<sub>1c</sub>) detection using ion-exchange high performance liquid chromatography (HPLC) method by clinically silent hemoglobin variant in University Malaya Medical Centre (UMMC)—A case report. *Clinical Biochemistry*, 42(4):430 – 434, 2009. ISSN 0009-9120. doi:[10.1016/j.clinbiochem.2008.10.015](https://doi.org/10.1016/j.clinbiochem.2008.10.015).
- [38] E. N. Lazareva, A. Y. Zyubin, I. G. Samusev, V. A. Slezhkin, V. I. Kochubey, and V. V. Tuchin. Refraction, fluorescence, and Raman spectroscopy of normal and glycated hemoglobin. In *Biophotonics: Photonic Solutions for Better Health Care VI*, volume 10685, pages 636 – 643. International Society for Optics and Photonics, SPIE, 2018. doi:[10.1117/12.2307102](https://doi.org/10.1117/12.2307102).
- [39] R. Pandey, N. C. Dingari, N. Spegazzini, R. R. Dasari, G. L. Horowitz, I. Barman. Emerging trends in optical sensing of glycemic markers for

- diabetes monitoring. *TrAC Trends in Analytical Chemistry*, 64:100 – 108, 2015. ISSN 0165-9936. doi:[10.1016/j.trac.2014.09.005](https://doi.org/10.1016/j.trac.2014.09.005).
- [40] R. Pandey. Raman Spectroscopy-Based sensing of glycated hemoglobin : Critical analysis and future outlook . *Journal of Postdoctoral Research*, 3(2): 8–16, 2015. URL [http://www.postdocjournal.com/file\\_journal/766\\_23669297.pdf](http://www.postdocjournal.com/file_journal/766_23669297.pdf).
- [41] M. Mallya, R. Shenoy, G. Kodyalamoole, M. Biswas, J. Karumathil, and S. Kamath. Absorption spectroscopy for the estimation of glycated hemoglobin (HbA<sub>1c</sub>) for the diagnosis and management of diabetes mellitus: A pilot study. *Photomedicine and Laser Surgery*, 31(5):219–224, 2013. doi:[10.1089/pho.2012.3421](https://doi.org/10.1089/pho.2012.3421).
- [42] O. S. Zhernovaya, A. N. Bashkatov, E. A. Genina, V. V. Tuchin, I. V. Meglinski, D. Yu. Churmakov, and L. J. Ritchie. Investigation of glucose-hemoglobin interaction by optical coherence tomography. In Valery V. Tuchin, editor, *Saratov Fall Meeting 2006: Optical Technologies in Biophysics and Medicine VIII*, volume 6535, pages 423 – 429. International Society for Optics and Photonics, SPIE, 2007. doi:[10.1117/12.740985](https://doi.org/10.1117/12.740985).
- [43] G. Mazarevica, T. Freivalds, and A. Jurka. Properties of erythrocyte light refraction in diabetic patients. *Journal of Biomedical Optics*, 7(2):244 – 247, 2002. doi:[10.1117/1.1463043](https://doi.org/10.1117/1.1463043).
- [44] O. A. Smolyanskaya, E. N. Lazareva, S. S. Nalegaev, N. V. Petrov, K. I. Zaytsev, P. A. Timoshina, D. K. Tuchina, Ya. G. Toropova, O. V. Korniyushin, A. Yu. Babenko, J.-P. Guillet, and V. V. Tuchin. Multimodal optical diagnostics of glycated biological tissues. *Biochemistry (Moscow)*, 84(1):124–143, Jan 2019. ISSN 1608-3040. doi:[10.1134/S0006297919140086](https://doi.org/10.1134/S0006297919140086).
- [45] T. Pan, M. Li, J. Chen, and H. Xue. Quantification of glycated hemoglobin indicator HbA<sub>1c</sub> through near-infrared spectroscopy. *Journal of Innovative Optical Health Sciences*, 07(4):1350060, 2014. doi:[10.1142/S1793545813500600](https://doi.org/10.1142/S1793545813500600).
- [46] I. Barman, N. C. Dingari, J. W. Kang, G. L. Horowitz, R. R. Dasari, and M. S. Feld. Raman spectroscopy-based sensitive and specific detection

- of glycated hemoglobin. *Analytical Chemistry*, 84(5):2474–2482, 2012. doi:[10.1021/ac203266a](https://doi.org/10.1021/ac203266a).
- [47] M. Syamala Kiran, T. Itoh, K.-i. Yoshida, N. Kawashima, V. Biju, and M. Ishikawa. Selective detection of HbA1c using surface enhanced resonance Raman spectroscopy. *Analytical Chemistry*, 82(4):1342–1348, 2010. doi:[10.1021/ac902364h](https://doi.org/10.1021/ac902364h).
- [48] J. Lin, L. Shao, S. Qiu, X. Huang, M. Liu, Z. Zheng, D. Lin, Y. Xu, Z. Li, Y. Lin, R. Chen, and S. Feng. Application of a near-infrared laser tweezers Raman spectroscopy system for label-free analysis and differentiation of diabetic red blood cells. *Biomed. Opt. Express*, 9(3):984–993, Mar 2018. doi:[10.1364/BOE.9.000984](https://doi.org/10.1364/BOE.9.000984).
- [49] R. Pandey, S. P. Singh, C. Zhang, G. L. Horowitz, N. Lue, L. Galindo, R. R. Dasari, and I. Barman. Label-free spectrochemical probe for determination of hemoglobin glycation in clinical blood samples. *Journal of Biophotonics*, 11(10):e201700397, 2018. doi:[10.1002/jbio.201700397](https://doi.org/10.1002/jbio.201700397).
- [50] J. F. Villa-Manríquez, J. Castro-Ramos, F. Gutiérrez-Delgado, M. A. López-Pacheco, and A. E. Villanueva-Luna. Raman spectroscopy and PCA-SVM as a non-invasive diagnostic tool to identify and classify qualitatively glycated hemoglobin levels in vivo. *Journal of Biophotonics*, 10(8):1074–1079, 2017. doi:[10.1002/jbio.201600169](https://doi.org/10.1002/jbio.201600169).
- [51] O. Cherkasova, M. Nazarov, and A. Shkurinov. Noninvasive blood glucose monitoring in the terahertz frequency range. *Optical and Quantum Electronics*, 48(3):217, 2016. doi:[10.1007/s11082-016-0490-5](https://doi.org/10.1007/s11082-016-0490-5).
- [52] D. L. Felipe, J. M. Hempe, S. Liu, N. Matter, J. Maynard, C. Linares, and S. A. Chalew. Skin intrinsic fluorescence is associated with Hemoglobin A1c and hemoglobin glycation index but not mean blood glucose in children with type 1 diabetes. *Diabetes Care*, 34(8):1816–1820, 2011. ISSN 0149-5992. doi:[10.2337/dc11-0049](https://doi.org/10.2337/dc11-0049).
- [53] S. Kim and J. Lee. Noninvasive apparatus and method for testing glycated hemoglobin, March 2016. URL <https://patents.google.com/patent/US20160061810A1/en>. visited on 30-12-2019.

- [54] R. H. Thompson. Calorimetric glucose oxidase method for blood glucose. *Clin. Chim. Acta*, 13(3):133–135, 1966. doi:[10.1016/0009-8981\(66\)90281-6](https://doi.org/10.1016/0009-8981(66)90281-6).
- [55] Wiener Laboratorios. Enzymatic glycemia AA, 2000. URL <https://www.wiener-lab.com.ar/VademecumDocumentos/Vademecumingles/glicemia{ }enzimatica{ }aa{ }en.pdf>.
- [56] L. Silveira Junior, Rita De C., Fernandes Borges, R. S. Navarro, H. E. Giana, R. A. Zângaro, M. Tadeu, T.s Pacheco, and A. B. Fernandes. Quantifying glucose and lipid components in human serum by Raman spectroscopy and multivariate statistics. *Lasers in Medical Science*, 2017. doi:[10.1007/s10103-017-2173-2](https://doi.org/10.1007/s10103-017-2173-2).
- [57] Rita De C., F. Borges, R. Scarparo Navarro, H.r E. Giana, F. G. Tavares, A. B. Fernandes, and L. Silveira Junior. Detecting alterations of glucose and lipid components in human serum by near-infrared Raman spectroscopy. *Research on Biomedical Engineering*, 31(2):160–168, 2015. doi:[10.1590/2446-4740.0593](https://doi.org/10.1590/2446-4740.0593).
- [58] J. Xue, H. Chen, D. Xiong, G. Huang, H. Ai, Y. Liang, and X. Yan. Noninvasive Measurement of Glucose in Artificial Plasma with Near-Infrared and Raman Spectroscopy. *Applied Optics*, 68(4):428–433, 2014. doi:[10.1366/13-07250](https://doi.org/10.1366/13-07250).
- [59] D. Rohleder, W. Petrich, D. Gmbh, and S. Str. Quantitative analysis of serum and serum ultrafiltrate by means of Raman spectroscopy. *The Analyst*, 129:906–911, 2004. doi:[10.1039/b408927h](https://doi.org/10.1039/b408927h).
- [60] D. Rohleder and K. Gerber. Comparison of mid-infrared and Raman spectroscopy in the quantitative analysis of serum. *Journal of Biomedical Optics*, 10(June):1–10, 2005. doi:[10.1117/1.1911847](https://doi.org/10.1117/1.1911847).
- [61] A. J. Berger, T.-w. Koo, I. Itzkan, G. Horowitz, and M. S. Feld. Multicomponent blood analysis by near-infrared Raman spectroscopy. *Applied Optics*, 38(13), 1999. doi:[10.1364/AO.38.002916](https://doi.org/10.1364/AO.38.002916).
- [62] J. Y. Qu, B. C. Wilson, and D. Suria. Concentration measurements of multiple analytes in human sera by near-infrared laser Raman spectroscopy. *Applied Optics*, 38(25), 1999. doi:[10.1364/AO.38.005491](https://doi.org/10.1364/AO.38.005491).

- [63] Parachalil, D. R., Bruno, C., Bonnier, F., Blasco, H., Chourpa, I., Baker, M. J., McIntyre, J., Byrne, H. J. Analysis of bodily fluids using vibrational spectroscopy: a direct comparison of raman scattering and infrared absorption techniques for the case of glucose in blood serum. *Analyst*, 144:3334–3346, 2019. <http://dx.doi.org/10.1039/C9AN00125E>.
- [64] Giansante, S., Giana, H. E., Fernandes, A. B., Silveira, L. Analytical performance of Raman spectroscopy in assaying biochemical components in human serum. *Lasers in Medical Science*, 2021. ISSN 1435-604X. <https://doi.org/10.1007/s10103-021-03247-8>.
- [65] J. Shao, M. Lin, Y. Li, X. Li, J. Liu, J. Liang, and H. Yao. In vivo blood glucose quantification using raman spectroscopy. *PLoS ONE*, 7(10):e48127, 2012. doi:[10.1371/journal.pone.0048127](https://doi.org/10.1371/journal.pone.0048127).
- [66] M. J. Scholtes-Timmerman, S. Bijlsma, M. J. Fokkert, R. Slingerland, and S. J. F. van Veen. Raman spectroscopy as a promising tool for noninvasive point-of-care glucose monitoring. *Journal of Diabetes Science and Technology*, 8(5):974–979, 2014. doi:[10.1177/1932296814543104](https://doi.org/10.1177/1932296814543104). PMID: 25037192.
- [67] S. M. Lundsgaard-Nielsen, A. Pors, S. O. Banke, J. E. Henriksen, D.h K. Hepp, and A. Weber. Critical-depth raman spectroscopy enables home-use non-invasive glucose monitoring. *PLOS ONE*, 13(5):1–11, 05 2018. doi:[10.1371/journal.pone.0197134](https://doi.org/10.1371/journal.pone.0197134).
- [68] K. Ma, Jonathan M. Yuen, N. C. Shah, J. T. Walsh, M. R. Glucksberg, and R. P. Van Duyne. In vivo, transcutaneous glucose sensing using surface-enhanced spatially offset raman spectroscopy: Multiple rats, improved hypoglycemic accuracy, low incident power, and continuous monitoring for greater than 17 days. *Analytical Chemistry*, 83(23): 9146–9152, 2011. doi:[10.1021/ac202343e](https://doi.org/10.1021/ac202343e).
- [69] J. Chaiken, W. Finney, P. E. Knudson, R. S. Weinstock, M. Khan, R. J. Bussjager, D. Hagrman, P. Hagrman, Y. Zhao, C. M. Peterson, and K. Peterson. Effect of hemoglobin concentration variation on the accuracy and precision of glucose analysis using tissue modulated, noninvasive, in

- vivo raman spectroscopy of human blood: a small clinical study. *Journal of Biomedical Optics*, 10(3), May 2005. ISSN 1083-3668. doi:[10.1117/1.1922147](https://doi.org/10.1117/1.1922147).
- [70] A. M. K. Enejder, T. G. Scecina, J. Oh, M. Hunter, W. Shih, S. Sasic, G. L. Horowitz, and M. S. Feld. Raman spectroscopy for noninvasive glucose measurements. *Journal of Biomedical Optics*, 10(3):1 – 9, 2005. doi:[10.1117/1.1920212](https://doi.org/10.1117/1.1920212).
- [71] D. A. Stuart, J. M. Yuen, N. Shah, O. Lyandres, C. R. Yonzon, M. R. Glucksberg, J. T. Walsh, and R. P. Van Duyne. In vivo glucose measurement by surface-enhanced raman spectroscopy. *Analytical Chemistry*, 78(20):7211–7215, 2006. doi:[10.1021/ac061238u](https://doi.org/10.1021/ac061238u). PMID: 17037923.
- [72] J. W. Kang, Y. S. Park, H. Chang, W. Lee, S. P. Singh, W. Choi, L. H. Galindo, R. R. Dasari, S. H. Nam, J. Park, and P. T. C. So. Direct observation of glucose fingerprint using in vivo raman spectroscopy. *Science Advances*, 6(4), 2020. doi:[10.1126/sciadv.aay5206](https://doi.org/10.1126/sciadv.aay5206).
- [73] Joseph Chaiken and Charles M. Peterson. Noninvasive blood and tissue analysis: Raman spectroscopy, one perspective for monitoring of glucose and beyond. *Journal of Diabetes Science and Technology*, 15(1): 28–33, 2021. doi:[10.1177/1932296820964803](https://doi.org/10.1177/1932296820964803). URL <https://doi.org/10.1177/1932296820964803>. PMID: 33084386.
- [74] Qiaoyun Wang, Guangfei Wu, Feifei Pian, Peng Shan, Zhigang Li, and Zhenhe Ma. Simultaneous detection of glucose, triglycerides, and total cholesterol in whole blood by fourier-transform raman spectroscopy. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 260:119906, 2021. ISSN 1386-1425. doi:<https://doi.org/10.1016/j.saa.2021.119906>. URL <https://www.sciencedirect.com/science/article/pii/S1386142521004820>.
- [75] Cheng-Ju Sung, Szu-Han Chao, and Shih-Chieh Hsu. Rapid detection of glucose on nanostructured gold film biosensor by surface-enhanced raman spectroscopy. *Biosensors*, 11(2), 2021. ISSN 2079-6374.

- doi:[10.3390/bios11020054](https://doi.org/10.3390/bios11020054). URL <https://www.mdpi.com/2079-6374/11/2/54>.
- [76] H. Michael Heise, Sven Delbeck, and Ralf Marbach. Noninvasive monitoring of glucose using near-infrared reflection spectroscopy of skin—constraints and effective novel strategy in multivariate calibration. *Biosensors*, 11(3), 2021. ISSN 2079-6374. doi:[10.3390/bios11030064](https://doi.org/10.3390/bios11030064). URL <https://www.mdpi.com/2079-6374/11/3/64>.
- [77] Maciej S. Wróbel, Jeong Hee Kim, Piyush Raj, Ishan Barman, and Janusz Smulko. Utilizing pulse dynamics for non-invasive raman spectroscopy of blood analytes. *Biosensors and Bioelectronics*, 180:113115, 2021. ISSN 0956-5663. doi:<https://doi.org/10.1016/j.bios.2021.113115>. URL <https://www.sciencedirect.com/science/article/pii/S0956566321001524>.
- [78] Z. Movasaghi, S. Rehman, and I. U. Rehman. Raman spectroscopy of biological tissues. *Applied Spectroscopy Reviews*, 42(5):493–541, 2007. doi:[10.1080/05704920701551530](https://doi.org/10.1080/05704920701551530).
- [79] J. M. Shaver. *Chemometrics for Raman Spectroscopy*. Taylor & Francis Group, New York, 2001. ISBN 0-8247-0557-2.
- [80] E. Smith and G. Dent. *The Theory of Raman Spectroscopy*. John Wiley & Sons, Ltd, 2005. ISBN 9780470011836. doi:[10.1002/0470011831.ch3](https://doi.org/10.1002/0470011831.ch3).
- [81] M. Schlösser. *Theory of Quantitative Raman Spectroscopy*. Springer International Publishing, Cham, 2014. doi:[10.1007/978-3-319-06221-1\\_3](https://doi.org/10.1007/978-3-319-06221-1_3).
- [82] C. V. Raman and K. S. Krishnan. A New Type of Secondary Radiation. *Nature*, 121(3048):501–502, 1928. ISSN 1476-4687. doi:[10.1038/121501c0](https://doi.org/10.1038/121501c0).
- [83] D. Lin-Vien, N. B. Colthup, W. G. Fateley, and J. G. Grasselli. Chapter 1 - introduction. In D. Lin-Vien, N. B. Colthup, W. G. Fateley, and J. G. Grasselli, editors, *The Handbook of Infrared and Raman Characteristic Frequencies of Organic Molecules*, pages 1 – 7. Academic Press, San Diego, 1991. ISBN 978-0-12-451160-6. doi:[10.1016/B978-0-08-057116-4.50007-9](https://doi.org/10.1016/B978-0-08-057116-4.50007-9).
- [84] J. R. Ferraro, K. Nakamoto, and C. W. Brown. Chapter 1 - basic theory. In J. R. Ferraro, K. Nakamoto, and C. W. Brown, editors,



- Introductory Raman Spectroscopy (Second Edition)*, pages 1 – 94. Academic Press, San Diego, second edition edition, 2003. ISBN 978-0-12-254105-6. doi:[10.1016/B978-012254105-6/50004-4](https://doi.org/10.1016/B978-012254105-6/50004-4).
- [85] G. Gliemann. K. Nakamoto: Infrared and raman spectra of inorganic and coordination compounds. *John Wiley and Sons, New York*, 82(11):1263–1263, 1978. doi:[10.1002/bbpc.19780821138](https://doi.org/10.1002/bbpc.19780821138).
- [86] H. Cartwright. Physical Chemistry. By Peter Atkins, Oxford University Press: Oxford, U.K. Student's Solutions Manual and Instructor's Solutions Manual are also available. *The Chemical Educator*, 6(4):262–263, 2001. ISSN 1430-4171. doi:[10.1007/s00897010493a](https://doi.org/10.1007/s00897010493a).
- [87] S. Krimm. The infrared spectra of complex molecules, vol. 1 (3rd ed.), I. J. Bellamy, Halsted Press, a division of John Wiley & Sons, Inc., New York, 1975. *Journal of Polymer Science: Polymer Letters Edition*, 14(2):121–121, 1976. doi:[10.1002/pol.1976.130140217](https://doi.org/10.1002/pol.1976.130140217).
- [88] H. Hu, J. Bai, G. Xia, W. Zhang, and Y. Ma. Improved Baseline Correction Method Based on Polynomial Fitting for Raman Spectroscopy. *Photonic Sensors*, 8(4):332–340, 2018. doi:[10.1007/s13320-018-0512-y](https://doi.org/10.1007/s13320-018-0512-y).
- [89] N. B. Colthup, L. H. Daly, and S. E. Wiberley. Chapter 1 - vibrational and rotational spectra. In Norman B. Colthup, Lawrence H. Daly, and Stephen E. Wiberley, editors, *Introduction to Infrared and Raman Spectroscopy (Third Edition)*, pages 1 – 73. Academic Press, San Diego, third edition edition, 1990. ISBN 978-0-12-182554-6. doi:[10.1016/B978-0-08-091740-5.50004-1](https://doi.org/10.1016/B978-0-08-091740-5.50004-1).
- [90] R. Sarpeshkar, T. Delbruck, and C. A. Mead. White noise in MOS transistors and resistors. *IEEE Circuits and Devices Magazine*, 9(6):23–29, Nov 1993. doi:[10.1109/101.261888](https://doi.org/10.1109/101.261888).
- [91] B. Schrader, editor. *Tools for infrared and Raman spectroscopy*, chapter 3, pages 63–188. John Wiley & Sons, Ltd, 1995. ISBN 9783527615438. doi:[10.1002/9783527615438.ch03](https://doi.org/10.1002/9783527615438.ch03).



- [92] R. L. McCreery. *Raman Spectroscopy for Chemical Analysis*. John Wiley & Sons, Inc., 2000. ISBN 0471252875.
- [93] A. Mahadevan-jansen and R. Richards-kortum. Raman spectroscopy for the detection of cancers and precancers. *Journal of Biomedical Optics*, 1(1): 31–70, 1996. doi:[10.1117/12.227815](https://doi.org/10.1117/12.227815).
- [94] M. León-bejarano, G. Dorantes-méndez, M. Ramírez-elías, M. O. Méndez, A. Alba, I. Rodríguez-leyva, and M. Jiménez. Fluorescence background removal method for biological Raman spectroscopy based on Empirical Mode Decomposition. *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2:3610–3613, 2016. doi:[10.1109/EMBC.2016.7591509](https://doi.org/10.1109/EMBC.2016.7591509).
- [95] A. E. Villanueva-Luna, J. Castro-Ramos, S. Vazquez-Montiel, A. Flores-Gil, J. A. Delgado-Atencio, and E. E. Orozco-Guillen. Fluorescence and noise subtraction from Raman spectra by using wavelets. *Optical Memory and Neural Networks*, 19(4):310–317, 2010. ISSN 1060-992X. doi:[10.3103/S1060992X10040089](https://doi.org/10.3103/S1060992X10040089).
- [96] R. Bro, N. D. Sidiropoulos, and A. K. Smilde. Maximum likelihood fitting using ordinary least squares. *Journal of Chemometrics*, 16:387–400, 2002. doi:[10.1002/cem.734](https://doi.org/10.1002/cem.734).
- [97] Člupek, M., Matějka, P. and Volka, K. Noise reduction in Raman spectra : Finite impulse response filtration versus Savitzky – Golay smoothing. *Journal of Raman Spectroscopy*, 38:1174–1179, 2007. doi:[10.1002/jrs](https://doi.org/10.1002/jrs).
- [98] H. N. Chavarría-Lizárraga. *Mapeo de tejido biológico utilizando espectroscopía Raman*. Master thesis, Instituto Nacional de Astrofísica Óptica y Electrónica, 2019.
- [99] A. E. Villanueva-Luna. *Espectroscopia Raman en fluidos biológicos extracelulares [Raman Spectroscopy in extra-cellular biological fluids]*. Phd thesis, Instituto Nacional de Astrofísica Óptica y Electrónica, 2013.
- [100] S. Shalev-Shwartz and S. Ben-David. *Understanding Machine Learning: From Theory to Algorithms*. Cambridge University Press, 2014. URL <https://www.cs.huji.ac.il/~shais/UnderstandingMachineLearning/>.

- [101] T. Dietterich. Overfitting and undercomputing in machine learning. *ACM Comput. Surv.*, 27(3):326–327, September 1995. ISSN 0360-0300. doi:[10.1145/212094.212114](https://doi.org/10.1145/212094.212114).
- [102] B. Ghogh and M. Crowley. The theory behind overfitting, cross validation, regularization, bagging, and boosting: Tutorial, 2019.
- [103] E. Guevara, J. C. Torres-Galván, M. G. Ramírez-Elías, C. Luevano-Contreras, and F. J. González. Use of Raman spectroscopy to screen diabetes mellitus with machine learning tools. *Biomed. Opt. Express*, 9(10):4998–5010, Oct 2018. doi:[10.1364/BOE.9.004998](https://doi.org/10.1364/BOE.9.004998).
- [104] B. Özbalci, I. H. Boyaci, A. Topcu, C. Kadilar, and U. Tamer. Rapid analysis of sugars in honey by processing Raman spectrum using chemometric methods and artificial neural networks. *Food Chemistry*, 136(3):1444–1452, 2013. doi:[10.1016/j.foodchem.2012.09.064](https://doi.org/10.1016/j.foodchem.2012.09.064).
- [105] L. Nørgaard, A. Saudland, J. Wagner, J. P. Nielsen, L. Munck, and S. B. Engelsen. Interval Partial Least-Squares Regression (iPLS): A comparative chemometric study with an example from Near-Infrared spectroscopy. *Applied Spectroscopy*, 54(3):413–419, 2000. doi:[10.1366/0003702001949500](https://doi.org/10.1366/0003702001949500).
- [106] E. Micheli-tzanakou. *Supervised and unsupervised pattern recognition: feature extraction and computational intelligence*. CRC Press, 1999. ISBN 0849322782.
- [107] P. Gómez-Gil. INAOE course: Topics Avanzados Redes Neuronales Artificiales: Red de Perceptrones de varios niveles, 2012. URL <https://ccc.inaoep.mx/~pgomez/cursos/redes%20neuronales%20artificiales/presentaciones/retropropagacion.pdf>.
- [108] M. Moller. A scaled conjugate gradient algorithm for fast supervised learning. neural networks 6, 525–533. *Neural Networks*, 6:525–533, 12 1993. doi:[10.1016/So893-6080\(05\)80056-5](https://doi.org/10.1016/So893-6080(05)80056-5).
- [109] MathWorks. Levenberg-marquardt backpropagation, 2020. URL <https://la.mathworks.com/help/deeplearning/ref/trainlm.html>. Accessed 21 October 2020.

- 
- [110] A. Landi, P. Piaggi, M. Laurino, and D. Menicucci. Artificial neural networks for nonlinear regression and classification. In *2010 10th International Conference on Intelligent Systems Design and Applications*, pages 115–120, Nov 2010. doi:[10.1109/ISDA.2010.5687280](https://doi.org/10.1109/ISDA.2010.5687280).
- [111] S. Holmes course, Stanford University Statistics. RMS Error, 2000. URL <http://statweb.stanford.edu/~susan/courses/s60/split/node60.html>.
- [112] R. O. Duda, D. G. Stork, and P. E. Hart. *Pattern Classification*. Wiley Interscience, 2001. ISBN 0-471-05669-3.
- [113] J. Brownlee. What is a Confusion Matrix in Machine Learning, 2016. URL <https://machinelearningmastery.com/confusion-matrix-machine-learning/>.
- [114] P. Pavlidis, I. Wapinski, and W. S. Noble. Support vector machine classification on the web. *Bioinformatics*, 20(4):586–587, 2004. ISSN 13674803. doi:[10.1093/bioinformatics/btg461](https://doi.org/10.1093/bioinformatics/btg461).
- [115] G. M. Fung and O. L. Mangasarian. Multicategory Proximal Support Vector Machine Classifiers. *Machine Learning*, 59(1):77–97, 2005. ISSN 1573-0565. doi:[10.1007/s10994-005-0463-6](https://doi.org/10.1007/s10994-005-0463-6).
- [116] J. F. Villa Manriquez. *Diagnostico de diabetes mediante espectroscopia Raman y hemoglobina glicosilada*. Phd thesis, Instituto Nacional de Astrofísica Óptica y Electrónica, 2017.
- [117] A. Ben-Hur, C. S. Ong, S. Sonnenburg, B. Schölkopf, and G. Rätsch. Support vector machines and kernels for computational biology. *PLOS Computational Biology*, 4(10):1–10, 10 2008. doi:[10.1371/journal.pcbi.1000173](https://doi.org/10.1371/journal.pcbi.1000173).
- [118] M. Peixeiro. Linear regression-understanding the theory, 2018. URL <https://towardsdatascience.com/linear-regression-understanding-the-theory-7e53ac2831b5>. Accessed 21 October 2020.

- 
- [119] J. Murdoch and J. A. Barnes. *Linear regression theory*. Palgrave Macmillan UK, London, 1973. ISBN 978-1-349-01063-9. doi:[10.1007/978-1-349-01063-9\\_9](https://doi.org/10.1007/978-1-349-01063-9_9).
  - [120] T. Kohonen. The Self-Organizing Map. *Proceedings of the IEEE*, 78(9), 1990. doi:[10.1109/5.58325](https://doi.org/10.1109/5.58325).
  - [121] J. R. Hilera and V. J. Martinez. *Redes neuronales artificiales : fundamentos , modelos y aplicaciones*. RA-MA, 1995. ISBN 8478971556.
  - [122] J. M. Marín Diazaraque. Los mapas auto-organizados de Kohonen (SOM), 2007. URL <http://halweb.uc3m.es/esp/Personal/personas/jmmarin/esp/DM/tema5dm.pdf>.
  - [123] B. M. Brío and A. S. Molina. *Redes Neuronales y Sistemas Borrosos. 3a Edición*. Textos Universitarios. RA-MA S.A. Editorial y Publicaciones, 2006. ISBN 9788478977437. URL <https://books.google.com.mx/books?id=KwbjGAAACAAJ>.
  - [124] I. Guyon and A. Elisseeff. *An Introduction to Feature Extraction*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2006. ISBN 978-3-540-35488-8. doi:[10.1007/978-3-540-35488-8\\_1](https://doi.org/10.1007/978-3-540-35488-8_1).
  - [125] S. Chatterjee. What is feature extraction? feature extraction in image processing, 2020. URL <https://www.mygreatlearning.com/blog/feature-extraction-in-image-processing/>. Accessed 21 October 2020.
  - [126] MathWorks. Feature extraction for machine learning and deep learning, 2020. URL <https://www.mathworks.com/discovery/feature-extraction.html#:~:text=%20Feature%20extraction%20can%20be%20accomplished%20manually%20or,to%20extract%20features%20automatically%20from%20signals...%20More%20>. Accessed 21 October 2020.
  - [127] L. I. Smith. A tutorial on Principal Components Analysis, 2002. URL [http://www.cs.otago.ac.nz/cosc453/student\\_tutorials/principal\\_components.pdf](http://www.cs.otago.ac.nz/cosc453/student_tutorials/principal_components.pdf).
-

- [128] J. Brownlee. How to choose a feature selection method for machine learning, 2019. URL <https://machinelearningmastery.com/feature-selection-with-real-and-categorical-data/>.
- [129] Z. Zhao, F. Morstatter, S. Sharma, S. Alelyani, A. Anand, and H. Liu. Advancing Feature Selection Research. *ASU Feature Selection Repository Arizona State University*, pages 1 – 28, 2010. URL [http://featureselection.asu.edu/featureselection\[\\_\]techreport.pdf](http://featureselection.asu.edu/featureselection[_]techreport.pdf).
- [130] M. Doshi and S. K Chaturvedi. Correlation Based Feature Selection (CFS) Technique to Predict Student Perfomance. *International journal of Computer Networks & Communications*, 6(3):197–206, 2014. ISSN 09752293. doi:[10.5121/ijcnc.2014.6315](https://doi.org/10.5121/ijcnc.2014.6315).
- [131] E. Morales course, INAOE. Búsqueda, 2000. URL <https://ccc.inaoep.mx/{~}emorales/Cursos/Aprendizaje2/Acetatos/busqueda2.pdf>.
- [132] K. Kira and L. A. Rendell. A practical approach to feature selection. In Derek Sleeman and Peter Edwards, editors, *Machine Learning Proceedings 1992*, pages 249–256. Morgan Kaufmann, San Francisco (CA), 1992. ISBN 978-1-55860-247-2. doi:[10.1016/B978-1-55860-247-2.50037-1](https://doi.org/10.1016/B978-1-55860-247-2.50037-1).
- [133] M. Robnik-Šikonja and I. Kononenko. An adaptation of {R}elief for attribute estimation in regression. *Machine {L}earning: {P}roceedings of the {F}ourteenth International Conference (ICML'97)*, 5:296–304, 1997. URL <http://lkm.fri.uni-lj.si/rmarko/papers/robnik97-icml.pdf>.
- [134] R. Kohavi and G. H. John. Wrappers for feature subset selection. *Artificial Intelligence*, 97(1):273 – 324, 1997. ISSN 0004-3702. doi:[10.1016/S0004-3702\(97\)00043-X](https://doi.org/10.1016/S0004-3702(97)00043-X).
- [135] R. Chang and K. A. Goldsby. *Chemistry*. McGraw-Hill Education, 11 edition, 2014.
- [136] N. González-Viveros, J. Castro-Ramos, P. Gómez-Gil, and H. H. Cerecedo-Núñez. On the estimation of sugars concentrations sing raman spectroscopy and artificial neural networks. *Food Chemistry*, page 119077, 2021. ISSN 1386-1425. doi:[10.1016/j.foodchem.2021.129375](https://doi.org/10.1016/j.foodchem.2021.129375).

- 
- [137] D. Tosi and G. Perrone. *Optical fiber sensors for biomedical applications*. Artech House, 2017. ISBN 978-1630811525.
- [138] C. Weykamp. HbA<sub>1c</sub>: A review of analytical and clinical aspects. *Ann. Lab. Med.*, 33(6):393–400, 2013. doi:[10.3343/alm.2013.33.6.393](https://doi.org/10.3343/alm.2013.33.6.393).
- [139] P. Flowers, K. Theopold, R. Langley, and W. R. Robinson. *Chemistry 2e*. OpenStax, 2015. ISBN 978-1-947172-61-6.
- [140] Water Science School, Science for a changing world. Water density, 2007. URL [https://www.usgs.gov/special-topic/water-science-school/science/water-density?qt-science\\_center\\_objects=0#qt-science\\_center\\_objects](https://www.usgs.gov/special-topic/water-science-school/science/water-density?qt-science_center_objects=0#qt-science_center_objects). Accessed at Jan. 28, 2021.
- [141] Spinreact México. Determinación cuantitativa de glucosa, ivd, 2020. URL [https://www.spinreact.com.mx/public/\\_pdf/1001190.pdf](https://www.spinreact.com.mx/public/_pdf/1001190.pdf). Accessed 23 October 2020.
- [142] Thorlabs, Inc. Raman spectroscopy: The approach, 2021. URL [https://www.thorlabs.com/newgrouppage9.cfm?objectgroup\\_id=9014](https://www.thorlabs.com/newgrouppage9.cfm?objectgroup_id=9014). Accessed 18 February 2021.
- [143] Laser Institute of America. *American National Standard for Safe Use of Lasers*. American National Standards, 2007. ISBN 978-1-940168-00-5.
- [144] S. Jensen, MIT engineering. How do glucometers work?, 2011. URL <https://engineering.mit.edu/engage/ask-an-engineer/how-do-glucometers-work/>. Accessed 28- Enero - 2021.
- [145] Wiener Lab group. método de inhibición inmunoturbidimétrica para la determinación cuantitativa de hba<sub>1c</sub>, 2020. URL [https://www.wiener-lab.com.ar/VademecumDocumentos/Vademecum{%}20espanol/hba1c\\_v2\\_turbitest\\_aa\\_sp.pdf](https://www.wiener-lab.com.ar/VademecumDocumentos/Vademecum{%}20espanol/hba1c_v2_turbitest_aa_sp.pdf). Accessed 23 October 2020.
- [146] M. H. Beale, T. M. Hagan, and H. B. Demuth. *Neural Network Toolbox™ 7 User's Guide*. MathWorks, 2010. ISBN 0971732108.
- [147] M. T. Hagan, H. B. Demuth, M. H. Beale. *Neural Network Design*. PWS Publishing, 1996. ISBN 978-0-534-94332-5.
-

- 
- [148] M. H. Beale, M. T. Hagan, H. B. Demuth. *Deep Learning Toolbox: User's Guide*. The MathWorks, Inc, 2021.
- [149] M.T. Hagan and M.B. Menhaj. Training feedforward networks with the marquardt algorithm. *IEEE Transactions on Neural Networks*, 5(6):989–993, 1994. doi:[10.1109/72.329697](https://doi.org/10.1109/72.329697).
- [150] Martin Foddslette Møller. A scaled conjugate gradient algorithm for fast supervised learning. *Neural Networks*, 6(4):525–533, 1993. ISSN 0893-6080. doi:[https://doi.org/10.1016/So893-6080\(05\)80056-5](https://doi.org/10.1016/So893-6080(05)80056-5). URL <https://www.sciencedirect.com/science/article/pii/S0893608005800565>.
- [151] Donald W. Marquardt. An Algorithm for Least-Squares Estimation of Nonlinear Parameters. *Journal of the Society for Industrial and Applied Mathematics*, 11(2):431–441, 1963. ISSN 03684245. URL <http://www.jstor.org/stable/2098941>. Publisher: Society for Industrial and Applied Mathematics.
- [152] Waitako New Zeland. weka 3: data mining with open source machine learning software in java, 2020. URL <https://www.cs.waikato.ac.nz/~ml/weka/>. Accessed 31 October 2020.
- [153] K. Hèberger. Chapter 7 - chemoinformatics—multivariate mathematical–statistical methods for data evaluation. In Károly Vékey, András Telekes, and Akos Vertes, editors, *Medical Applications of Mass Spectrometry*, pages 141–169. Elsevier, Amsterdam, 2008. ISBN 978-0-444-51980-1. doi:[10.1016/B978-044451980-1.50009-4](https://doi.org/10.1016/B978-044451980-1.50009-4).
- [154] I. H. Boyaci, H. T. Temiz, H. E. Geniş, E. Acar Soykut, N. N. Yazgan, B. Güven, R. S. Uysal, A. G. Bozkurt, K. İlaslan, O. Torun, F. C. Dudak Şeker. Dispersive and FT-Raman spectroscopic methods in food analysis. *RSC Adv.*, 5(70):56606–56624, 2015. doi:[10.1039/C4RA12463D](https://doi.org/10.1039/C4RA12463D).
- [155] W. L. Clarke. The Original Clarke Error Grid Analysis (EGA). *Diabetes technology and therapeutics*, 7(5):776–779, 2005. doi:[10.1089/dia.2005.7.776](https://doi.org/10.1089/dia.2005.7.776).
- [156] Barry H. Ginsberg. Factors affecting blood glucose monitoring: Sources of errors in measurement. *Journal of Diabetes Science and Technology*, 3(4): 903–913, 2009. doi:[10.1177/193229680900300438](https://doi.org/10.1177/193229680900300438). PMID: 20144340.



- [157] Ksenia Tonyushkina and James H. Nichols. Glucose meters: A review of technical challenges to obtaining accurate results. *Journal of Diabetes Science and Technology*, 3(4):971–980, 2009. doi:[10.1177/193229680900300446](https://doi.org/10.1177/193229680900300446). PMID: 20144348.
- [158] X. Qiu, H. Huang, Z. Huang, Z. Zhuang, Z.i Guo, and S. Liu. Effect of red light-emitting diodes irradiation on hemoglobin for potential hypertension treatment based on confocal Micro-Raman spectroscopy. *Scanning*, page 5067867, 2017. doi:[10.1155/2017/5067867](https://doi.org/10.1155/2017/5067867).
- [159] C. G. Atkins, K. Buckley, M. W. Blades, and R. F. B. Turner. Raman spectroscopy of blood and blood components. *Applied Spectroscopy*, 71(5):767–793, 2017. doi:[10.1177/0003702816686593](https://doi.org/10.1177/0003702816686593).
- [160] B. R. Wood, B. Tait, and D. McNaughton. Micro-Raman characterisation of the R to T state transition of haemoglobin within a single living erythrocyte. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, 1539(1):58 – 70, 2001. ISSN 0167-4889. doi:[10.1016/S0167-4889\(01\)00089-1](https://doi.org/10.1016/S0167-4889(01)00089-1).
- [161] J. Lin, J. Lin, Z. Huang, P. Lu, J. Wang, X. Wang, R. Chen. Raman spectroscopy of human hemoglobin for diabetes detection. *Journal of Innovative Optical Health Sciences*, 07(01):1350051, 2014. doi:[10.1142/S179354581350051X](https://doi.org/10.1142/S179354581350051X).
- [162] H. Sato, H. Chiba, H. Tashiro, and Y. Ozaki. Excitation wavelength-dependent changes in Raman spectra of whole blood and hemoglobin : comparison of the spectra with 514.5-, 720-, and 1064-nm excitation. *Journal of Biomedical Optics*, 6(3):366–370, 2001. doi:[10.1117/1.1380668](https://doi.org/10.1117/1.1380668).
- [163] T. Makhnii, O. Ilchenko, A. Reynt, Y. Pilgun, A. Kutsyk, D. Krasnenkov, M. Ivasyuk, and V. Kukharsky. Age-related changes in FTIR and Raman spectra of human blood. *Ukrainian Journal of Physics*, 61(10):853–862, 2016. ISSN 20710194. doi:[10.15407/ujpe61.10.0853](https://doi.org/10.15407/ujpe61.10.0853).
- [164] A. M. Pawlak, J. R. Beattie, J. V. Glenn, A. W. Stitt, and J. J. McGarvey. Raman spectroscopy of advanced glycation end products (AGEs),

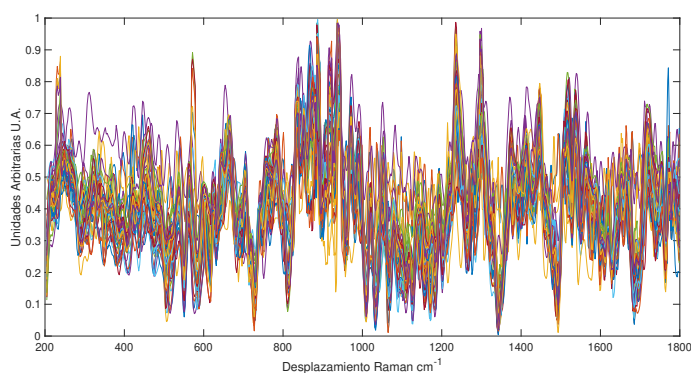


- possible markers for progressive retinal dysfunction. *Journal of Raman Spectroscopy*, 39(11):1635–1642, 2008. doi:[10.1002/jrs.2011](https://doi.org/10.1002/jrs.2011).
- [165] J. De Gelder, K. De Gussem, P. Vandenabeele, and L. Moens. Reference database of raman spectra of biological molecules. *Journal of Raman Spectroscopy*, 38(9):1133–1147, 2007. doi:[10.1002/jrs.1734](https://doi.org/10.1002/jrs.1734).
- [166] Semantic Scholar. Detection of glucose levels using excitation and difference raman spectroscopy at the iusl. In *New York State Center for Advanced Technology in Photonics Applications at City University of New York*, 2004.
- [167] k. Ilaslan, I. H Boyaci, and A. Topcu. Rapid analysis of glucose, fructose and sucrose contents of commercial soft drinks using Raman spectroscopy. *Food Control*, 48:56 – 61, 2015. doi:[10.1016/j.foodcont.2014.01.001](https://doi.org/10.1016/j.foodcont.2014.01.001).
- [168] GT lab, Argentina. determinación de glucosa en plasma, suero u orina, 2020. URL [https://www.gtlab.com.ar/UserFiles/mediaManager/1/b180b50d7d709df5b606a3a72988ad86fdeabae4\\_cf9dc0b65e16472e907b45147bbbaa8a12a6c947.pdf](https://www.gtlab.com.ar/UserFiles/mediaManager/1/b180b50d7d709df5b606a3a72988ad86fdeabae4_cf9dc0b65e16472e907b45147bbbaa8a12a6c947.pdf). Accessed 21 Mayo 2021.
- [169] María-Eugenia Torregrosa, Juan Molina, Carlos R. Argente, and Javier Ena. Evaluación de tres sistemas de determinación rápida de hemoglobina a1c para monitorización del control glucémico en pacientes con diabetes mellitus. 62(10):478–484, 2015. ISSN 1575-0922. doi:[10.1016/j.endonu.2015.07.007](https://doi.org/10.1016/j.endonu.2015.07.007). Publisher: Elsevier.
- [170] Sujaytha Paknikar, Rohan Sarmah, Losika Sivaganeshan, Adam Welke, Al Rizzo, Kirk Larson, and Marc Rendell. Long-term performance of point-of-care hemoglobin a1c assays. *Journal of Diabetes Science and Technology*, 10(6):1308–1315, 2016. doi:[10.1177/1932296816645362](https://doi.org/10.1177/1932296816645362). PMID: 27113451.

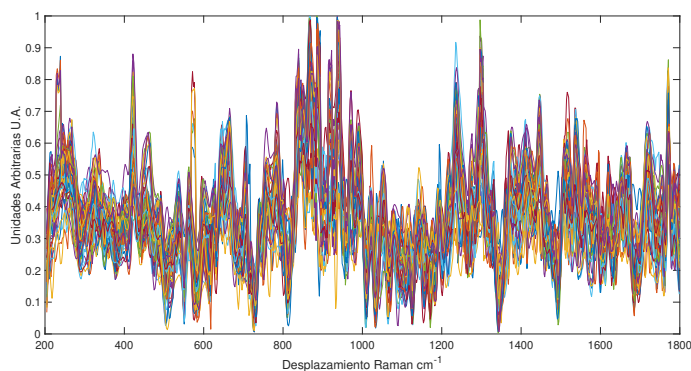


## FIGURAS Y TABLAS COMPLEMENTARIAS

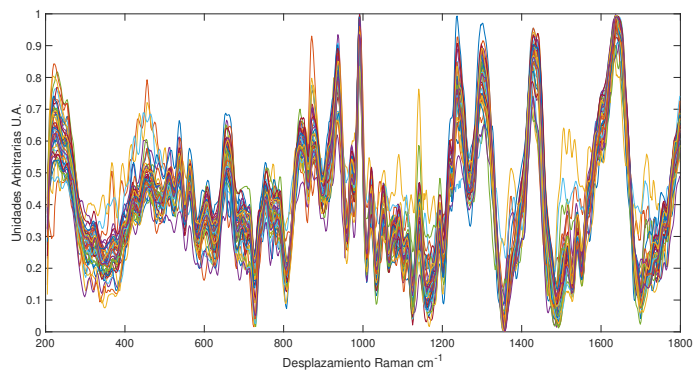
### A.1. Espectros de mediciones *in vivo*



**Figura A.1:** Espectros Raman promedio por voluntario en la región de muñeca.

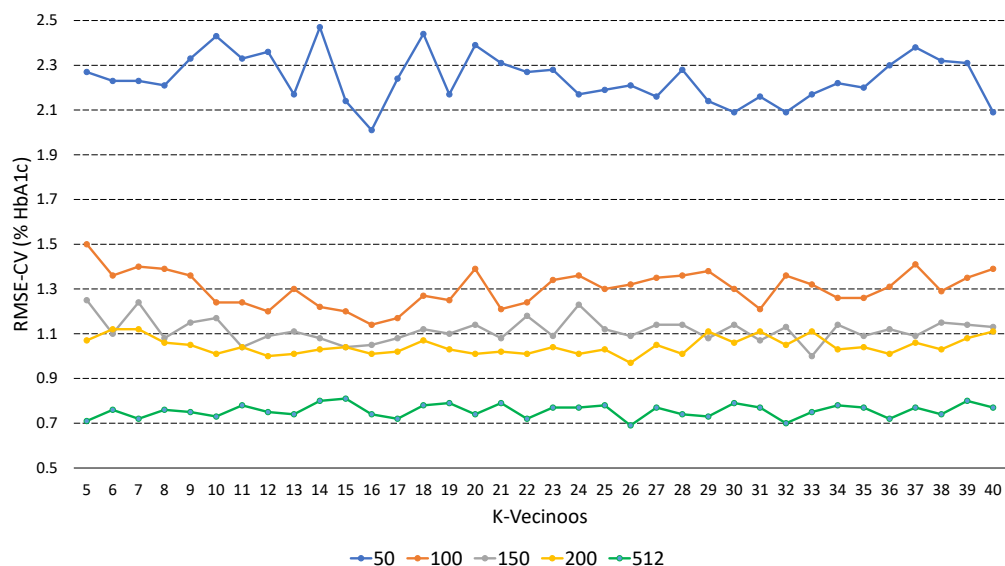


**Figura A.2:** Espectros Raman promedio por voluntario en la región de dedo.

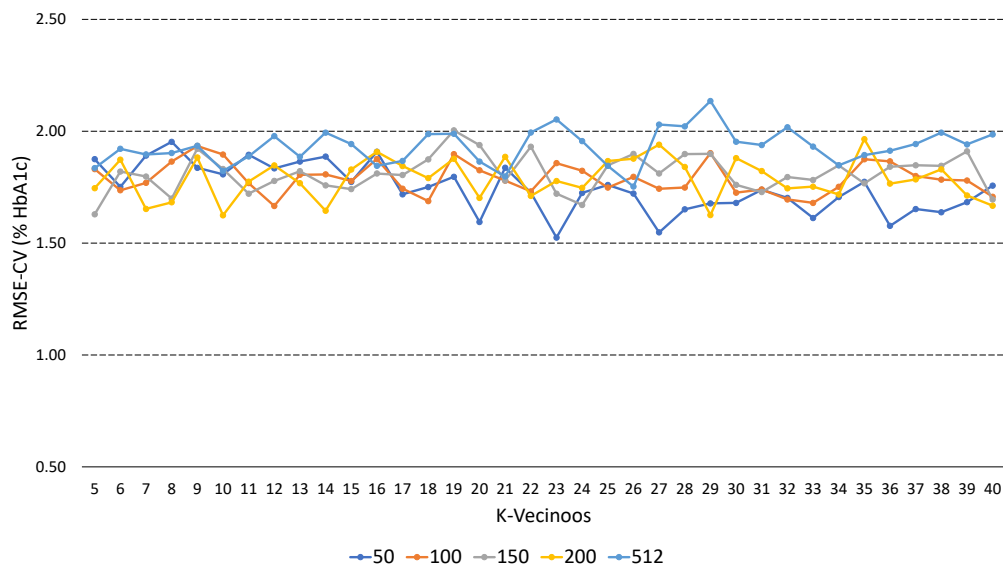


**Figura A.3:** Espectros Raman promedio por voluntario en suero.

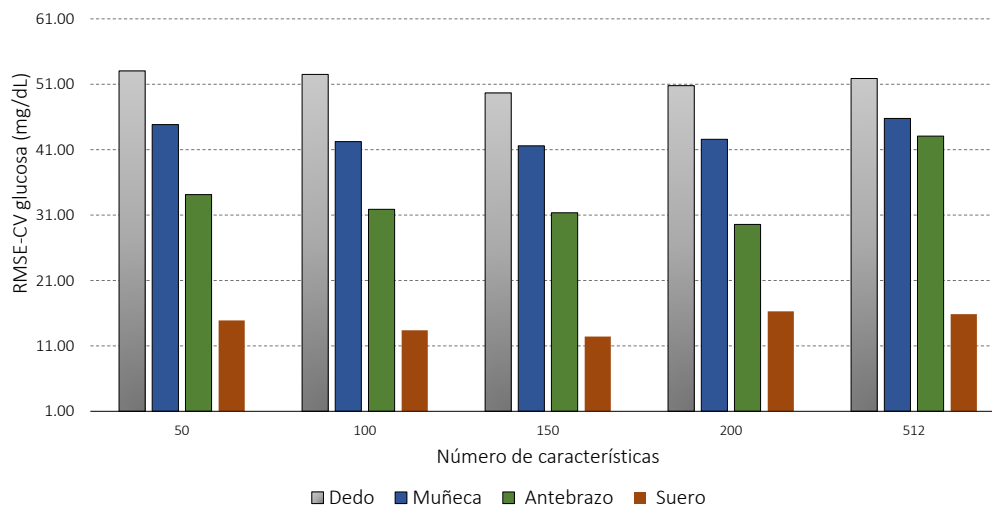
## A.2. Análisis de k-vecinos y características por región



**Figura A.4:** Variación de K-Vecinos por número de características en región de la muñeca.

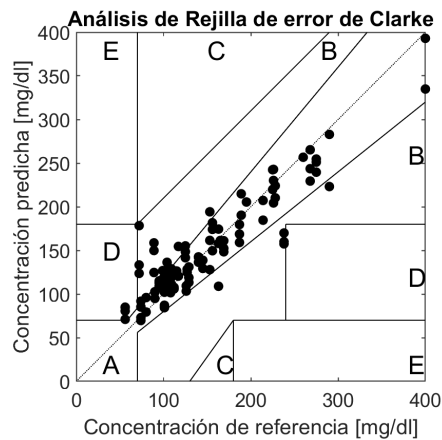


**Figura A.5:** Variación de K-Vecinos por número de características en región del dedo.

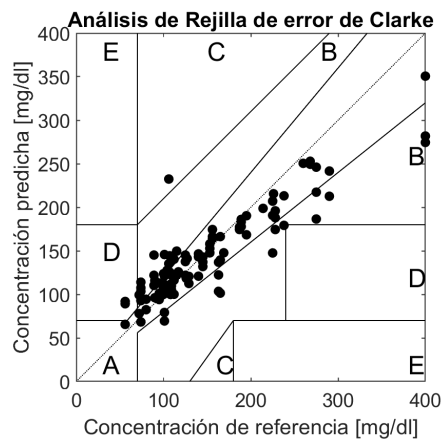


**Figura A.6:** RMSE-CV predicción de glucosa por parte del cuerpo y número de características usando RReliefF.

### A.3. Rejilla de error de Clarke por región

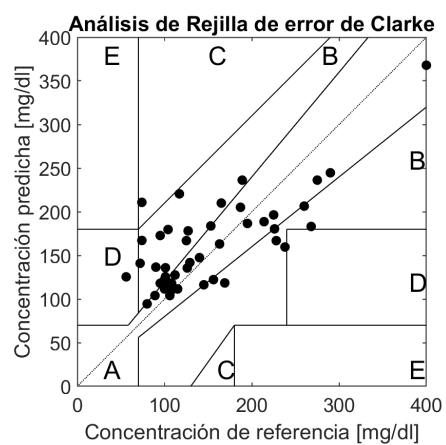


**Figura A.7:** Rejilla de error de Clarke para las predicción de glucosa en antebrazo.



**Figura A.8:** Rejilla de error de Clarke para las predicción de glucosa en muñeca.

#### A.4. Análisis espectral mediciones *in vivo*



**Figura A.9:** Rejilla de error de Clarke para las predicción de glucosa en dedo.

Tabla A.1: Picos localizados de HbA1c en muñeca.

| %    | 1578    | 1571    | 1536   | 1436    | 1311    | 1308    | 1230    | 1222    | 1114    | 1106    | 969    | 749    | 665   |
|------|---------|---------|--------|---------|---------|---------|---------|---------|---------|---------|--------|--------|-------|
| 5.2  | 1590±12 | 1565±6  | 1539±3 | 1447±11 | 1331±20 | 1302±6  | 1240±10 | 1218±4  | 1134±20 | 1100±6  | 972±3  | 740±9  | 663±2 |
| 5.3  | 1578±0  | 1563±8  | 1531±5 | 1437±1  | 1324±13 | 1307±1  | 1228±2  | 1215±7  | 1113±1  | 1099±7  | 958±11 | 754±5  | 668±3 |
| 5.4  | 1589±11 | 1564±7  | 1537±1 | 1448±12 | 1331±20 | 1299±9  | 1237±7  | 1219±3  | 1113±1  | 1097±9  | 971±2  | 738±11 | 663±2 |
| 5.5  | 1590±12 | 1563±8  | 1540±4 | 1447±11 | 1328±17 | 1301±7  | 1240±10 | 1218±4  | 1119±5  | 1099±7  | 972±3  | 740±9  | 662±3 |
| 6    | 1588±10 | 1564±7  | 1538±2 | 1444±8  | 1298±13 | 1298±10 | 1237±7  | 1193±29 | 1116±2  | 1098±8  | 972±3  | 761±12 | 666±1 |
| 6.1  | 1589±11 | 1565±6  | 1539±3 | 1448±12 | 1330±19 | 1301±7  | 1240±10 | 1218±4  | 1122±8  | 1101±5  | 972±3  | 741±8  | 664±1 |
| 6.2  | 1589±11 | 1565±6  | 1539±3 | 1446±10 | 1328±17 | 1328±20 | 1238±8  | 1218±4  | 1118±4  | 1098±8  | 972±3  | 743±6  | 663±2 |
| 6.3  | 1579±1  | 1563±8  | 1540±4 | 1447±11 | 1319±8  | 1301±7  | 1235±5  | 1217±5  | 1115±1  | 1097±9  | 973±4  | 741±8  | 657±8 |
| 6.4  | 1589±11 | 1564±7  | 1537±1 | 1446±10 | 1326±15 | 1326±18 | 1237±7  | 1220±2  | 1115±1  | 1098±8  | 972±3  | 765±16 | 665±0 |
| 6.5  | 1579±1  | 1561±10 | 1538±2 | 1446±10 | 1329±18 | 1300±8  | 1238±8  | 1195±27 | 1119±5  | 1100±6  | 972±3  | 742±7  | 664±1 |
| 6.6  | 1590±12 | 1564±7  | 1539±3 | 1447±11 | 1298±13 | 1298±10 | 1237±7  | 1219±3  | 1114±0  | 1076±30 | 972±3  | 739±10 | 662±3 |
| 6.7  | 1592±14 | 1573±2  | 1540±4 | 1449±13 | 1300±11 | 1300±8  | 1237±7  | 1197±25 | 1119±5  | 1092±14 | 973±4  | 757±8  | 663±2 |
| 6.8  | 1590±12 | 1564±7  | 1537±1 | 1447±11 | 1298±13 | 1298±10 | 1237±7  | 1218±4  | 1114±0  | 1097±9  | 972±3  | 739±10 | 664±1 |
| 7.3  | 1590±12 | 1561±10 | 1537±1 | 1447±11 | 1313±2  | 1299±9  | 1235±5  | 1208±14 | 1112±2  | 1112±6  | 971±2  | 741±8  | 665±0 |
| 7.4  | 1580±2  | 1562±9  | 1538±2 | 1446±10 | 1328±17 | 1300±8  | 1237±7  | 1237±15 | 1119±5  | 1119±13 | 972±3  | 756±7  | 662±3 |
| 7.8  | 1587±9  | 1565±6  | 1538±2 | 1446±10 | 1324±13 | 1299±9  | 1237±7  | 1237±15 | 1117±3  | 1098±8  | 972±3  | 766±17 | 666±1 |
| 7.9  | 1588±10 | 1563±8  | 1539±3 | 1446±10 | 1299±12 | 1299±9  | 1237±7  | 1237±15 | 1116±2  | 1116±10 | 972±3  | 758±9  | 665±0 |
| 8    | 1589±11 | 1561±10 | 1537±1 | 1446±10 | 1311±0  | 1298±10 | 1235±5  | 1207±15 | 1114±0  | 1086±20 | 968±1  | 742±7  | 668±3 |
| 8.1  | 1590±12 | 1566±5  | 1538±2 | 1447±11 | 1331±20 | 1301±7  | 1238±8  | 1218±4  | 1121±7  | 1099±7  | 972±3  | 740±9  | 664±1 |
| 8.3  | 1589±11 | 1564±7  | 1538±2 | 1445±9  | 1299±12 | 1299±9  | 1237±7  | 1221±1  | 1145±31 | 1080±26 | 972±3  | 759±10 | 664±1 |
| 8.6  | 1575±3  | 1563±8  | 1536±0 | 1446±10 | 1333±22 | 1300±8  | 1237±7  | 1194±28 | 1116±2  | 1096±10 | 971±2  | 743±6  | 659±6 |
| 9    | 1574±4  | 1561±10 | 1540±4 | 1448±12 | 1300±11 | 1300±8  | 1238±8  | 1197±25 | 1118±4  | 1094±12 | 972±3  | 718±31 | 663±2 |
| 9.3  | 1588±10 | 1561±10 | 1538±2 | 1447±11 | 1312±1  | 1299±9  | 1235±5  | 1195±27 | 1111±3  | 1078±28 | 970±1  | 743±6  | 664±1 |
| 9.8  | 1590±12 | 1565±6  | 1538±2 | 1449±13 | 1330±19 | 1301±7  | 1237±7  | 1217±5  | 1119±5  | 1099±7  | 972±3  | 758±9  | 663±2 |
| 10   | 1590±12 | 1560±11 | 1538±2 | 1446±10 | 1329±18 | 1298±10 | 1236±6  | 1193±29 | 1114±0  | 1076±30 | 972±3  | 741±8  | 666±1 |
| 10.1 | 1579±1  | 1562±9  | 1538±2 | 1446±10 | 1328±17 | 1300±8  | 1236±6  | 1195±27 | 1118±4  | 1081±25 | 972±3  | 757±8  | 660±5 |
| 10.2 | 1586±8  | 1567±4  | 1539±3 | 1443±7  | 1298±13 | 1298±10 | 1237±7  | 1196±26 | 1117±3  | 1082±24 | 972±3  | 745±4  | 665±0 |
| 11.1 | 1581±3  | 1562±9  | 1539±3 | 1446±10 | 1328±17 | 1300±8  | 1237±7  | 1194±28 | 1118±4  | 1081±25 | 972±3  | 758±9  | 662±3 |
| 11.2 | 1587±9  | 1565±6  | 1538±2 | 1446±10 | 1375±64 | 1299±9  | 1238±8  | 1220±2  | 1116±2  | 1080±26 | 972±3  | 758±9  | 665±0 |
| 11.8 | 1588±10 | 1564±7  | 1539±3 | 1446±10 | 1328±17 | 1300±8  | 1236±6  | 1195±27 | 1117±3  | 1081±25 | 972±3  | 743±6  | 661±4 |
| 12.1 | 1590±12 | 1562±9  | 1538±2 | 1447±11 | 1328±17 | 1300±8  | 1236±6  | 1195±27 | 1118±4  | 1080±26 | 971±2  | 743±6  | 663±2 |
| 12.2 | 1589±11 | 1560±11 | 1537±1 | 1447±11 | 1316±5  | 1299±9  | 1235±5  | 1207±15 | 1115±1  | 1080±26 | 987±18 | 741±8  | 665±0 |
| 12.7 | 1590±12 | 1565±6  | 1539±3 | 1447±11 | 1329±18 | 1300±8  | 1237±7  | 1195±27 | 1117±3  | 1092±14 | 972±3  | 741±8  | 662±3 |
| 13.4 | 1590±12 | 1559±12 | 1539±3 | 1447±11 | 1328±17 | 1299±9  | 1236±6  | 1221±1  | 1115±1  | 1098±8  | 972±3  | 741±8  | 666±1 |
| 13.8 | 1589±11 | 1560±11 | 1537±1 | 1447±11 | 1313±2  | 1299±9  | 1235±5  | 1195±27 | 1115±1  | 1079±27 | 971±2  | 740±9  | 665±0 |
| 14   | 1590±12 | 1562±9  | 1538±2 | 1447±11 | 1298±13 | 1298±10 | 1236±6  | 1195±27 | 1115±1  | 1079±27 | 971±2  | 740±9  | 664±1 |



Tabla A.2: Picos localizados de HbA<sub>1c</sub> en antebrazo.

| %    | 1578    | 1571    | 1536   | 1436    | 1311    | 1308    | 1230    | 1222    | 1114    | 1106    | 969   | 749    | 665    |
|------|---------|---------|--------|---------|---------|---------|---------|---------|---------|---------|-------|--------|--------|
| 5.2  | 1589±11 | 1566±5  | 1540±4 | 1434±2  | 1329±18 | 1303±5  | 1240±10 | 1216±6  | 1119±5  | 1098±8  | 973±4 | 760±11 | 663±2  |
| 5.3  | 1576±2  | 1564±7  | 1545±9 | 1437±1  | 1324±13 | 1293±15 | 1232±2  | 1216±6  | 1110±4  | 1098±8  | 972±3 | 740±9  | 664±1  |
| 5.4  | 1589±11 | 1565±6  | 1538±2 | 1447±11 | 1332±21 | 1297±11 | 1236±6  | 1220±2  | 1114±0  | 1097±9  | 972±3 | 739±10 | 662±3  |
| 5.5  | 1590±12 | 1562±9  | 1540±4 | 1446±10 | 1330±19 | 1301±7  | 1240±10 | 1216±6  | 1133±19 | 1100±6  | 973±4 | 742±7  | 662±3  |
| 6    | 1564±14 | 1564±7  | 1539±3 | 1444±8  | 1298±13 | 1298±10 | 1238±8  | 1221±1  | 1114±0  | 1079±27 | 972±3 | 759±10 | 664±1  |
| 6.1  | 1589±11 | 1564±7  | 1540±4 | 1447±11 | 1330±19 | 1301±7  | 1239±9  | 1217±5  | 1123±9  | 1097±9  | 973±4 | 741±8  | 662±3  |
| 6.2  | 1589±11 | 1565±6  | 1539±3 | 1446±10 | 1331±20 | 1299±9  | 1237±7  | 1219±3  | 1118±4  | 1097±9  | 972±3 | 767±18 | 662±3  |
| 6.3  | 1579±1  | 1563±8  | 1538±2 | 1446±10 | 1335±24 | 1297±11 | 1236±6  | 1218±4  | 1116±2  | 1097±9  | 971±2 | 740±9  | 663±2  |
| 6.4  | 1588±10 | 1563±8  | 1537±1 | 1446±10 | 1317±6  | 1298±10 | 1237±7  | 1237±15 | 1114±0  | 1099±7  | 972±3 | 763±14 | 664±1  |
| 6.5  | 1579±1  | 1562±9  | 1538±2 | 1446±10 | 1328±17 | 1299±9  | 1237±7  | 1237±15 | 1117±3  | 1117±11 | 971±2 | 742±7  | 661±4  |
| 6.6  | 1589±11 | 1564±7  | 1539±3 | 1447±11 | 1296±15 | 1296±12 | 1237±7  | 1219±3  | 1114±0  | 1097±9  | 972±3 | 756±7  | 662±3  |
| 6.7  | 1573±5  | 1562±9  | 1540±4 | 1448±12 | 1300±11 | 1300±8  | 1238±8  | 1221±1  | 1117±3  | 1093±13 | 972±3 | 758±9  | 664±1  |
| 6.8  | 1590±12 | 1564±7  | 1536±0 | 1447±11 | 1297±14 | 1297±11 | 1237±7  | 1221±1  | 1113±1  | 1098±8  | 971±2 | 740±9  | 663±2  |
| 7.3  | 1589±11 | 1563±8  | 1538±2 | 1446±10 | 1297±14 | 1297±11 | 1236±6  | 1236±14 | 1114±0  | 1090±16 | 971±2 | 740±9  | 664±1  |
| 7.4  | 1580±2  | 1562±9  | 1538±2 | 1445±9  | 1327±16 | 1299±9  | 1235±5  | 1235±13 | 1121±7  | 1100±6  | 972±3 | 757±8  | 661±4  |
| 7.8  | 1585±7  | 1563±8  | 1539±3 | 1445±9  | 1323±12 | 1299±9  | 1238±8  | 1219±3  | 1117±3  | 1102±4  | 971±2 | 758±9  | 665±0  |
| 7.9  | 1589±11 | 1564±7  | 1539±3 | 1446±10 | 1330±19 | 1299±9  | 1238±8  | 1216±6  | 1116±2  | 1116±10 | 972±3 | 743±6  | 666±1  |
| 8    | 1589±11 | 1563±8  | 1538±2 | 1447±11 | 1311±0  | 1297±11 | 1235±5  | 1235±13 | 1113±1  | 1113±7  | 969±0 | 740±9  | 655±10 |
| 8.1  | 1590±12 | 1564±7  | 1539±3 | 1448±12 | 1302±9  | 1302±6  | 1238±8  | 1216±6  | 1123±9  | 1102±4  | 974±5 | 741±8  | 661±4  |
| 8.3  | 1587±9  | 1563±8  | 1539±3 | 1445±9  | 1299±12 | 1299±9  | 1237±7  | 1237±15 | 1116±2  | 1116±10 | 972±3 | 760±11 | 664±1  |
| 8.6  | 1576±2  | 1563±8  | 1537±1 | 1445±9  | 1299±12 | 1299±9  | 1237±7  | 1221±1  | 1116±2  | 1095±11 | 970±1 | 760±11 | 660±5  |
| 9    | 1591±13 | 1572±1  | 1540±4 | 1448±12 | 1300±11 | 1300±8  | 1237±7  | 1237±15 | 1118±4  | 1118±12 | 971±2 | 787±38 | 665±0  |
| 9.3  | 1589±11 | 1562±9  | 1538±2 | 1446±10 | 1298±13 | 1298±10 | 1236±6  | 1236±14 | 1112±2  | 1112±6  | 972±3 | 741±8  | 663±2  |
| 9.8  | 1589±11 | 1564±7  | 1539±3 | 1449±13 | 1328±17 | 1300±8  | 1239±9  | 1219±3  | 1121±7  | 1101±5  | 972±3 | 740±9  | 663±2  |
| 10   | 1590±12 | 1561±10 | 1539±3 | 1446±10 | 1329±18 | 1298±10 | 1236±6  | 1220±2  | 1114±0  | 1094±12 | 972±3 | 740±9  | 663±2  |
| 10.1 | 1579±1  | 1563±8  | 1538±2 | 1446±10 | 1328±17 | 1299±9  | 1237±7  | 1237±15 | 1118±4  | 1118±12 | 971±2 | 757±8  | 661±4  |
| 10.2 | 1587±9  | 1568±3  | 1540±4 | 1440±4  | 1297±14 | 1297±11 | 1238±8  | 1238±16 | 1116±2  | 1116±10 | 972±3 | 763±14 | 662±3  |
| 11.1 | 1563±15 | 1563±8  | 1539±3 | 1447±11 | 1300±11 | 1300±8  | 1236±6  | 1219±3  | 1118±4  | 1094±12 | 972±3 | 758±9  | 660±5  |
| 11.2 | 1564±14 | 1564±7  | 1539±3 | 1446±10 | 1299±12 | 1299±9  | 1238±8  | 1218±4  | 1116±2  | 1116±10 | 972±3 | 759±10 | 664±1  |
| 11.8 | 1589±11 | 1563±8  | 1538±2 | 1446±10 | 1299±12 | 1299±9  | 1237±7  | 1237±15 | 1117±3  | 1117±11 | 972±3 | 757±8  | 662±3  |
| 12.1 | 1590±12 | 1563±8  | 1538±2 | 1447±11 | 1300±11 | 1300±8  | 1237±7  | 1237±15 | 1119±5  | 1119±13 | 972±3 | 757±8  | 664±1  |
| 12.2 | 1590±12 | 1561±10 | 1537±1 | 1446±10 | 1328±17 | 1299±9  | 1236±6  | 1211±11 | 1113±1  | 1113±7  | 970±1 | 743±6  | 664±1  |
| 12.7 | 1590±12 | 1563±8  | 1539±3 | 1446±10 | 1329±18 | 1300±8  | 1238±8  | 1238±16 | 1119±5  | 1092±14 | 972±3 | 742±7  | 661±4  |
| 13.4 | 1590±12 | 1560±11 | 1539±3 | 1446±10 | 1328±17 | 1298±10 | 1236±6  | 1236±14 | 1115±1  | 1115±9  | 972±3 | 740±9  | 668±3  |
| 13.8 | 1589±11 | 1563±8  | 1538±2 | 1447±11 | 1328±17 | 1298±10 | 1236±6  | 1236±14 | 1111±3  | 1111±5  | 971±2 | 740±9  | 663±2  |
| 14   | 1590±12 | 1563±8  | 1539±3 | 1447±11 | 1298±13 | 1298±10 | 1236±6  | 1236±14 | 1113±1  | 1113±7  | 971±2 | 740±9  | 660±5  |

Tabla A.3: Picos localizados de HbA1c en dedo.

| %    | 1578    | 1571    | 1536   | 1436    | 1311    | 1308    | 1230    | 1222    | 1114    | 1106    | 969    | 749    | 665    |
|------|---------|---------|--------|---------|---------|---------|---------|---------|---------|---------|--------|--------|--------|
| 5.2  | 1589±11 | 1563±8  | 1540±4 | 1448±12 | 1330±19 | 1301±7  | 1239±9  | 1217±5  | 1123±9  | 1098±8  | 973±4  | 741±8  | 662±3  |
| 5.3  | 1595±17 | 1567±4  | 1535±1 | 144±4   | 1326±15 | 1305±3  | 1244±14 | 1217±5  | 1114±0  | 1099±7  | 960±9  | 754±5  | 667±2  |
| 5.4  | 1577±1  | 1564±7  | 1537±1 | 1447±11 | 1334±23 | 1298±10 | 1237±7  | 1219±3  | 1114±0  | 1098±8  | 972±3  | 740±9  | 664±1  |
| 5.5  | 1590±12 | 1563±8  | 1539±3 | 1445±9  | 1328±17 | 1302±6  | 1238±8  | 1217±5  | 1109±5  | 1096±10 | 972±3  | 741±8  | 663±2  |
| 6    | 1583±5  | 1563±8  | 1538±2 | 1445±9  | 1299±12 | 1299±9  | 1237±7  | 1219±3  | 1114±0  | 1114±8  | 972±3  | 759±10 | 664±1  |
| 6.1  | 1590±12 | 1565±6  | 1539±3 | 1447±11 | 1330±19 | 1302±6  | 1239±9  | 1217±5  | 1122±8  | 1100±6  | 974±5  | 742±7  | 663±2  |
| 6.2  | 1590±12 | 1564±7  | 1539±3 | 1447±11 | 1300±11 | 1300±8  | 1237±7  | 1219±3  | 1115±1  | 1096±10 | 971±2  | 760±11 | 664±1  |
| 6.3  | 1580±2  | 1563±8  | 1537±1 | 1446±10 | 1333±22 | 1301±7  | 1236±6  | 1218±4  | 1115±1  | 1097±9  | 971±2  | 742±7  | 662±3  |
| 6.4  | 1590±12 | 1563±8  | 1538±2 | 1447±11 | 1298±13 | 1298±10 | 1237±7  | 1221±1  | 1114±0  | 1099±7  | 972±3  | 741±8  | 663±2  |
| 6.5  | 1590±12 | 1562±9  | 1538±2 | 1446±10 | 1328±17 | 1300±8  | 1237±7  | 1237±15 | 1118±4  | 1118±12 | 971±2  | 742±7  | 662±3  |
| 6.6  | 1590±12 | 1564±7  | 1537±1 | 1434±2  | 1312±1  | 1299±9  | 1235±5  | 1219±3  | 1100±14 | 1100±6  | 972±3  | 739±10 | 665±0  |
| 6.7  | 1589±11 | 1571±0  | 1539±3 | 1448±12 | 1313±2  | 1299±9  | 1238±8  | 1219±3  | 1117±3  | 1093±13 | 972±3  | 763±14 | 664±1  |
| 6.8  | 1591±13 | 1562±9  | 1537±1 | 1448±12 | 1311±0  | 1298±10 | 1236±6  | 1236±14 | 1113±1  | 1098±8  | 971±2  | 740±9  | 664±1  |
| 7.3  | 1589±11 | 1560±11 | 1536±0 | 1446±10 | 1326±15 | 1299±9  | 1236±6  | 1236±14 | 1114±0  | 1114±8  | 941±28 | 742±7  | 666±1  |
| 7.4  | 1579±1  | 1563±8  | 1538±2 | 1446±10 | 1327±16 | 1299±9  | 1236±6  | 1236±14 | 1119±5  | 1119±13 | 972±3  | 741±8  | 661±4  |
| 7.8  | 1588±10 | 1565±6  | 1538±2 | 1446±10 | 1324±13 | 1300±8  | 1237±7  | 1237±15 | 1116±2  | 1098±8  | 972±3  | 758±9  | 665±0  |
| 7.9  | 1588±10 | 1565±6  | 1536±0 | 1446±10 | 1323±12 | 1299±9  | 1237±7  | 1218±4  | 1117±3  | 1103±3  | 972±3  | 757±8  | 666±1  |
| 8    | 1589±11 | 1559±12 | 1535±1 | 1446±10 | 1314±3  | 1298±10 | 1235±5  | 1206±16 | 1113±1  | 1113±7  | 997±28 | 742±7  | 648±17 |
| 8.1  | 1590±12 | 1564±7  | 1538±2 | 1428±8  | 1329±18 | 1302±6  | 1239±9  | 1215±7  | 1122±8  | 1101±5  | 972±3  | 741±8  | 663±2  |
| 8.3  | 1590±12 | 1564±7  | 1539±3 | 1446±10 | 1299±12 | 1299±9  | 1236±6  | 1220±2  | 1117±3  | 1085±21 | 972±3  | 761±12 | 665±0  |
| 8.6  | 1576±2  | 1563±8  | 1537±1 | 1444±8  | 1299±12 | 1299±9  | 1237±7  | 1237±15 | 1114±0  | 1095±11 | 971±2  | 759±10 | 660±5  |
| 9    | 1576±2  | 1560±11 | 1539±3 | 1448±12 | 1300±11 | 1300±8  | 1238±8  | 1238±16 | 1118±4  | 1099±7  | 967±2  | 763±14 | 667±2  |
| 9.3  | 1590±12 | 1560±11 | 1535±1 | 1446±10 | 1316±5  | 1298±10 | 1234±4  | 1234±12 | 1113±1  | 1113±7  | 942±27 | 740±9  | 666±1  |
| 9.8  | 1590±12 | 1563±8  | 1538±2 | 1448±12 | 1301±10 | 1301±7  | 1238±8  | 1218±4  | 1121±7  | 1099±7  | 972±3  | 742±7  | 662±3  |
| 10   | 1590±12 | 1560±11 | 1538±2 | 1445±9  | 1328±17 | 1299±9  | 1235±5  | 1235±13 | 1115±1  | 1115±9  | 972±3  | 741±8  | 665±0  |
| 10.1 | 1591±13 | 1564±7  | 1538±2 | 1446±10 | 1328±17 | 1301±7  | 1237±7  | 1237±15 | 1116±2  | 1105±1  | 972±3  | 755±6  | 662±3  |
| 10.2 | 1587±9  | 1567±4  | 1539±3 | 1444±8  | 1299±12 | 1299±9  | 1237±7  | 1219±3  | 1116±2  | 1116±10 | 972±3  | 747±2  | 663±2  |
| 11.1 | 1590±12 | 1562±9  | 1539±3 | 1447±11 | 1329±18 | 1300±8  | 1237±7  | 1237±15 | 1119±5  | 1119±13 | 972±3  | 757±8  | 662±3  |
| 11.2 | 1589±11 | 1563±8  | 1538±2 | 1445±9  | 1299±12 | 1299±9  | 1237±7  | 1237±15 | 1116±2  | 1101±5  | 971±2  | 758±9  | 665±0  |
| 11.8 | 1577±1  | 1562±9  | 1538±2 | 1447±11 | 1328±17 | 1300±8  | 1236±6  | 1236±14 | 1120±6  | 1120±14 | 972±3  | 741±8  | 664±1  |
| 12.1 | 1589±11 | 1561±10 | 1538±2 | 1447±11 | 1328±17 | 1300±8  | 1236±6  | 1236±14 | 1118±4  | 1118±12 | 973±4  | 742±7  | 662±3  |
| 12.2 | 1592±14 | 1561±10 | 1537±1 | 1447±11 | 1313±2  | 1299±9  | 1235±5  | 1235±13 | 1117±3  | 1104±2  | 970±1  | 742±7  | 662±3  |
| 12.7 | 1591±13 | 1564±7  | 1539±3 | 1447±11 | 1330±19 | 1300±8  | 1237±7  | 1237±15 | 1120±6  | 1120±14 | 971±2  | 742±7  | 662±3  |
| 13.4 | 1591±13 | 1558±13 | 1539±3 | 1446±10 | 1329±18 | 1299±9  | 1236±6  | 1236±14 | 1116±2  | 1116±10 | 973±4  | 740±9  | 666±1  |
| 13.8 | 1588±10 | 1561±10 | 1538±2 | 1446±10 | 1311±0  | 1299±9  | 1236±6  | 1220±2  | 1112±2  | 1112±6  | 972±3  | 740±9  | 663±2  |
| 14   | 1589±11 | 1561±10 | 1537±1 | 1448±12 | 1299±12 | 1299±9  | 1236±6  | 1236±14 | 1113±1  | 1113±7  | 971±2  | 740±9  | 661±4  |

Tabla A.4: Picos localizados de glucosa en antebrazo.

| mg/dL | 405    | 439    | 544   | 769    | 837   | 908    | 1060    | 1106    | 1353    |
|-------|--------|--------|-------|--------|-------|--------|---------|---------|---------|
| 56    | 419±14 | 457±18 | 536±7 | 756±13 | 838±1 | 918±11 | 1054±6  | 1116±10 | 1374±21 |
| 72    | 422±17 | 443±5  | 536±8 | 768±1  | 837±0 | 910±2  | 1053±7  | 1116±10 | 1376±23 |
| 74    | 399±5  | 423±16 | 535±8 | 757±13 | 837±0 | 919±11 | 1051±9  | 1112±6  | 1363±10 |
| 80    | 395±10 | 453±14 | 536±8 | 758±11 | 840±4 | 889±19 | 1052±7  | 1116±10 | 1373±20 |
| 89    | 423±18 | 437±2  | 535±9 | 766±4  | 838±1 | 921±13 | 1052±7  | 1113±7  | 1378±25 |
| 90    | 419±14 | 453±15 | 536±7 | 762±7  | 839±2 | 906±2  | 1052±7  | 1100±6  | 1372±19 |
| 95    | 402±3  | 420±18 | 537±6 | 763±6  | 838±1 | 923±15 | 1051±8  | 1117±11 | 1372±19 |
| 99    | 396±8  | 421±18 | 536±7 | 761±8  | 837±0 | 917±10 | 1055±5  | 1115±9  | 1373±21 |
| 100   | 421±16 | 448±9  | 535±9 | 768±1  | 839±3 | 908±0  | 1052±7  | 1114±8  | 1376±23 |
| 101   | 404±1  | 451±12 | 537±6 | 763±6  | 838±1 | 920±13 | 1054±6  | 1123±17 | 1372±20 |
| 104   | 399±6  | 461±22 | 536±8 | 757±12 | 836±0 | 917±9  | 1050±10 | 1100±6  | 1372±20 |
| 106   | 396±9  | 430±9  | 535±8 | 761±8  | 836±0 | 919±11 | 1052±8  | 1112±7  | 1374±21 |
| 107   | 384±21 | 449±10 | 536±8 | 783±14 | 837±0 | 919±11 | 1052±7  | 1115±9  | 1376±23 |
| 108   | 405±0  | 451±12 | 539±4 | 760±9  | 839±2 | 908±1  | 1053±7  | 1098±8  | 1373±21 |
| 109   | 385±20 | 421±18 | 536±8 | 767±2  | 837±0 | 907±1  | 1052±8  | 1113±7  | 1364±12 |
| 112   | 392±13 | 445±7  | 540±3 | 762±7  | 842±5 | 906±2  | 1058±2  | 1116±10 | 1354±1  |
| 115   | 396±9  | 428±11 | 536±8 | 757±12 | 836±0 | 918±10 | 1051±9  | 1117±11 | 1372±19 |
| 117   | 394±10 | 429±9  | 536±8 | 764±5  | 836±1 | 918±10 | 1050±9  | 1117±11 | 1373±20 |
| 125   | 399±6  | 449±11 | 535±8 | 769±0  | 845±9 | 919±11 | 1052±8  | 1116±10 | 1374±22 |
| 126   | 415±11 | 451±12 | 537±6 | 760±9  | 838±1 | 888±20 | 1051±9  | 1102±4  | 1372±20 |
| 127   | 399±6  | 449±11 | 535±8 | 769±0  | 845±9 | 919±11 | 1052±8  | 1116±10 | 1374±22 |
| 129   | 391±13 | 445±6  | 540±3 | 762±7  | 843±6 | 905±3  | 1053±6  | 1117±11 | 1375±22 |
| 140   | 395±10 | 452±13 | 537±7 | 770±1  | 836±0 | 921±13 | 1050±9  | 1113±7  | 1347±6  |
| 145   | 395±10 | 424±15 | 535±8 | 760±9  | 836±1 | 919±11 | 1053±6  | 1100±5  | 1378±26 |
| 153   | 396±9  | 426±12 | 536±7 | 758±11 | 836±1 | 917±9  | 1053±7  | 1102±4  | 1376±24 |
| 156   | 394±10 | 425±14 | 537±7 | 759±10 | 837±0 | 918±10 | 1053±7  | 1114±8  | 1374±21 |
| 163   | 405±0  | 427±12 | 535±9 | 787±18 | 836±1 | 918±10 | 1052±7  | 1118±13 | 1382±30 |
| 165   | 395±10 | 425±14 | 535±8 | 758±12 | 837±0 | 918±10 | 1051±9  | 1118±12 | 1372±20 |
| 169   | 399±6  | 428±10 | 536±8 | 757±13 | 836±1 | 916±8  | 1050±9  | 1118±12 | 1373±20 |
| 187   | 396±9  | 423±16 | 537±7 | 756±13 | 835±2 | 918±10 | 1052±8  | 1116±10 | 1373±20 |
| 189   | 397±8  | 426±13 | 536±8 | 760±9  | 836±1 | 919±11 | 1051±9  | 1112±6  | 1372±20 |
| 195   | 396±9  | 457±18 | 536±8 | 763±6  | 836±1 | 919±11 | 1051±9  | 1111±5  | 1373±20 |
| 214   | 390±15 | 426±13 | 536±8 | 763±6  | 836±0 | 917±9  | 1055±4  | 1115±9  | 1379±26 |
| 225   | 397±7  | 430±9  | 536±8 | 757±12 | 835±2 | 919±11 | 1050±10 | 1119±13 | 1373±20 |
| 226   | 395±10 | 426±12 | 536±8 | 785±16 | 836±1 | 918±10 | 1051±9  | 1119±13 | 1372±20 |
| 228   | 397±8  | 424±14 | 536±8 | 759±10 | 836±0 | 919±11 | 1053±6  | 1116±10 | 1376±23 |
| 238   | 394±11 | 425±13 | 536±7 | 760±9  | 836±1 | 919±11 | 1053±7  | 1116±10 | 1376±24 |
| 260   | 392±13 | 424±14 | 536±8 | 759±10 | 836±0 | 919±11 | 1053±7  | 1116±10 | 1375±22 |
| 268   | 397±8  | 430±9  | 536±7 | 757±12 | 836±1 | 918±10 | 1050±10 | 1119±13 | 1372±19 |
| 275   | 395±10 | 421±18 | 537±7 | 762±7  | 836±1 | 919±11 | 1051±9  | 1113±7  | 1373±2  |
| 290   | 411±6  | 451±12 | 536±8 | 759±10 | 837±1 | 906±2  | 1052±8  | 1101±5  | 1373±21 |
| 400   | 394±11 | 423±16 | 537±7 | 762±7  | 836±1 | 919±11 | 1051±9  | 1111±6  | 1374±21 |

Tabla A.5: Picos localizados de glucosa en muñeca.

| mg/dL | 405    | 439    | 544   | 769    | 837   | 908    | 1060    | 1106    | 1353    |
|-------|--------|--------|-------|--------|-------|--------|---------|---------|---------|
| 56    | 399±6  | 420±19 | 536±7 | 758±11 | 842±5 | 919±11 | 1053±7  | 1114±8  | 1355±3  |
| 72    | 386±19 | 441±2  | 535±8 | 766±3  | 837±1 | 920±12 | 1052±8  | 1115±9  | 1376±24 |
| 74    | 411±6  | 430±9  | 534±9 | 776±7  | 835±1 | 919±12 | 1064±4  | 1116±10 | 1362±9  |
| 80    | 394±11 | 454±15 | 535±8 | 758±11 | 841±4 | 938±30 | 1053±7  | 1117±11 | 1372±20 |
| 89    | 385±19 | 442±3  | 536±8 | 767±2  | 837±1 | 919±12 | 1052±8  | 1115±9  | 1363±10 |
| 90    | 413±9  | 454±15 | 535±9 | 763±6  | 838±1 | 908±0  | 1051±9  | 1119±13 | 1372±19 |
| 95    | 419±14 | 451±13 | 536±8 | 764±5  | 838±1 | 923±15 | 1051±9  | 1117±12 | 1372±19 |
| 99    | 426±21 | 457±18 | 534±9 | 742±27 | 837±0 | 918±10 | 1054±6  | 1114±9  | 1374±22 |
| 100   | 421±16 | 421±18 | 536±8 | 785±15 | 839±2 | 911±4  | 1052±8  | 1114±8  | 1375±22 |
| 101   | 384±21 | 417±22 | 535±8 | 766±4  | 837±0 | 908±0  | 1053±7  | 1118±12 | 1373±20 |
| 104   | 396±9  | 428±11 | 536±8 | 756±13 | 837±0 | 918±11 | 1051±9  | 1119±13 | 1372±20 |
| 106   | 393±12 | 428±11 | 535±8 | 764±5  | 835±1 | 919±11 | 1050±9  | 1111±5  | 1374±22 |
| 107   | 383±22 | 449±10 | 536±8 | 765±4  | 837±0 | 920±12 | 1052±8  | 1114±8  | 1376±23 |
| 108   | 413±8  | 451±12 | 536±7 | 761±8  | 838±2 | 912±4  | 1050±10 | 1100±6  | 1372±20 |
| 109   | 385±20 | 421±17 | 535±9 | 765±4  | 839±3 | 908±0  | 1052±8  | 1114±8  | 1376±23 |
| 112   | 397±8  | 444±6  | 541±3 | 763±6  | 844±7 | 905±3  | 1079±20 | 1113±8  | 1352±1  |
| 115   | 397±8  | 427±12 | 536±8 | 758±11 | 836±1 | 918±10 | 1051±8  | 1117±11 | 1373±20 |
| 117   | 395±10 | 431±8  | 536±8 | 763±6  | 836±1 | 918±11 | 1050±10 | 1119±13 | 1372±20 |
| 125   | 400±5  | 448±9  | 534±9 | 770±1  | 845±9 | 936±28 | 1051±8  | 1117±11 | 1372±19 |
| 126   | 410±5  | 450±11 | 538±5 | 760±9  | 838±1 | 919±12 | 1077±17 | 1099±7  | 1371±18 |
| 127   | 395±10 | 422±17 | 537±6 | 764±5  | 837±0 | 918±10 | 1054±5  | 1115±9  | 1362±10 |
| 129   | 395±10 | 424±15 | 542±2 | 762±7  | 843±6 | 905±3  | 1078±18 | 1112±6  | 1352±1  |
| 140   | 396±9  | 421±18 | 535±9 | 765±4  | 837±0 | 920±12 | 1051±9  | 1114±8  | 1372±19 |
| 145   | 395±10 | 424±14 | 537±6 | 761±9  | 836±1 | 919±11 | 1053±7  | 1116±10 | 1377±24 |
| 153   | 394±11 | 425±14 | 537±7 | 766±3  | 836±0 | 918±11 | 1054±6  | 1117±11 | 1377±25 |
| 156   | 395±10 | 426±12 | 536±7 | 761±8  | 836±1 | 918±10 | 1054±6  | 1116±10 | 1375±23 |
| 163   | 405±0  | 427±12 | 535±9 | 785±16 | 836±1 | 918±10 | 1051±8  | 1118±12 | 1383±31 |
| 165   | 394±11 | 427±11 | 536±8 | 758±11 | 837±0 | 918±10 | 1050±10 | 1118±12 | 1372±19 |
| 169   | 397±8  | 426±12 | 535±8 | 757±12 | 837±0 | 916±8  | 1050±10 | 1118±13 | 1372±19 |
| 187   | 395±10 | 423±15 | 537±7 | 758±11 | 836±1 | 919±11 | 1053±7  | 1116±10 | 1375±23 |
| 189   | 394±11 | 423±15 | 536±8 | 762±7  | 836±1 | 919±11 | 1050±9  | 1113±7  | 1373±21 |
| 195   | 401±4  | 432±7  | 541±3 | 763±6  | 834±2 | 919±11 | 1078±18 | 1113±7  | 1344±8  |
| 214   | 426±22 | 426±12 | 537±6 | 759±10 | 836±1 | 918±10 | 1055±5  | 1115±9  | 1374±22 |
| 225   | 396±8  | 428±11 | 536±8 | 757±12 | 836±1 | 918±10 | 1051±9  | 1118±12 | 1373±20 |
| 226   | 394±11 | 427±12 | 536±7 | 763±6  | 837±0 | 918±10 | 1051±9  | 1117±11 | 1372±20 |
| 228   | 395±10 | 422±17 | 537±7 | 779±10 | 837±0 | 919±11 | 1053±7  | 1116±10 | 1375±23 |
| 238   | 394±11 | 428±11 | 535±8 | 759±10 | 836±1 | 919±11 | 1053±7  | 1116±10 | 1376±24 |
| 260   | 395±10 | 424±15 | 536±7 | 760±9  | 836±1 | 918±11 | 1054±6  | 1117±12 | 1377±24 |
| 268   | 396±8  | 428±10 | 536±8 | 758±11 | 836±1 | 918±10 | 1049±10 | 1118±12 | 1373±20 |
| 275   | 396±8  | 451±13 | 536±8 | 760±9  | 836±1 | 920±12 | 1051±9  | 1115±9  | 1373±20 |
| 290   | 419±14 | 419±19 | 535±8 | 758±11 | 837±0 | 903±5  | 1057±3  | 1119±13 | 1372±19 |
| 400   | 395±10 | 432±7  | 536±8 | 761±8  | 836±1 | 919±11 | 1051±9  | 1115±9  | 1346±6  |

Tabla A.6: Picos localizados de glucosa en dedo.

| mg/dL | 405    | 439    | 544    | 769    | 837    | 908    | 1060    | 1106    | 1353    |
|-------|--------|--------|--------|--------|--------|--------|---------|---------|---------|
| 56    | 398±7  | 419±20 | 535±9  | 758±11 | 840±3  | 919±11 | 1053±7  | 1136±31 | 1355±3  |
| 72    | 389±16 | 446±7  | 535±9  | 766±3  | 843±7  | 920±12 | 1051±9  | 1115±9  | 1378±25 |
| 74    | 369±36 | 422±17 | 536±8  | 755±14 | 839±2  | 918±11 | 1050±10 | 1116±11 | 1363±10 |
| 80    | 397±8  | 419±19 | 536±8  | 757±12 | 843±7  | 920±12 | 1052±8  | 1117±11 | 1355±2  |
| 89    | 385±20 | 442±3  | 537±7  | 784±15 | 841±4  | 920±12 | 1052±8  | 1113±7  | 1380±27 |
| 90    | 403±2  | 453±14 | 539±4  | 760±10 | 837±0  | 909±1  | 1056±4  | 1109±3  | 1373±20 |
| 95    | 404±1  | 422±17 | 564±20 | 759±1  | 840±3  | 923±15 | 1051±9  | 1114±8  | 1374±22 |
| 99    | 421±16 | 458±19 | 538±6  | 758±11 | 839±2  | 918±11 | 1053±7  | 1115±9  | 1372±19 |
| 100   | 365±40 | 421±18 | 536±7  | 766±3  | 843±6  | 908±1  | 1051±9  | 1100±6  | 1364±12 |
| 101   | 422±17 | 460±21 | 536±7  | 765±4  | 841±4  | 919±11 | 1051±8  | 1115±9  | 1377±24 |
| 104   | 399±6  | 424±14 | 536±8  | 760±10 | 838±1  | 918±10 | 1049±11 | 1119±13 | 1370±18 |
| 106   | 401±3  | 450±11 | 533±10 | 762±7  | 837±1  | 918±11 | 1050±10 | 1113±7  | 1345±7  |
| 107   | 420±15 | 448±10 | 534±10 | 762±7  | 840±3  | 906±2  | 1051±9  | 1115±9  | 1361±8  |
| 108   | 403±2  | 451±12 | 540±4  | 762±7  | 838±1  | 908±0  | 1051±9  | 1123±17 | 1362±10 |
| 109   | 421±16 | 441±3  | 537±7  | 766±3  | 843±6  | 908±0  | 1052±7  | 1113±7  | 1377±24 |
| 112   | 396±9  | 442±3  | 541±2  | 762±7  | 844±7  | 904±4  | 1062±2  | 1112±6  | 1352±0  |
| 115   | 397±8  | 424±14 | 537±7  | 758±11 | 837±0  | 918±10 | 1050±10 | 1120±14 | 1373±20 |
| 117   | 396±8  | 427±11 | 535±9  | 763±6  | 837±0  | 918±10 | 1050±10 | 1118±12 | 1373±20 |
| 125   | 403±2  | 447±8  | 535±9  | 771±1  | 845±8  | 919±11 | 1051±9  | 1116±11 | 1344±9  |
| 126   | 406±1  | 449±10 | 535±8  | 762±7  | 838±1  | 919±11 | 1057±3  | 1101±5  | 1372±19 |
| 127   | 402±3  | 458±19 | 537±6  | 760±9  | 840±3  | 918±10 | 1056±4  | 1115±9  | 1362±10 |
| 129   | 400±5  | 446±8  | 540±3  | 762±7  | 843±7  | 904±3  | 1067±7  | 1139±34 | 1352±1  |
| 140   | 403±2  | 422±16 | 535±9  | 765±4  | 843±6  | 919±11 | 1050±9  | 1113±7  | 1345±7  |
| 145   | 396±9  | 459±20 | 536±8  | 758±11 | 837±0  | 919±11 | 1053±7  | 1100±5  | 1377±25 |
| 153   | 372±33 | 422±16 | 535±9  | 758±11 | 839±2  | 903±5  | 1053±7  | 1116±10 | 1376±24 |
| 156   | 394±11 | 422±16 | 536±7  | 759±11 | 837±0  | 920±12 | 1053±7  | 1114±8  | 1377±24 |
| 163   | 368±37 | 425±14 | 534±10 | 763±7  | 836±1  | 918±10 | 1052±8  | 1118±13 | 1359±7  |
| 165   | 396±9  | 423±16 | 536±8  | 757±12 | 838±1  | 918±10 | 1050±10 | 1119±13 | 1372±20 |
| 169   | 398±7  | 425±14 | 537±7  | 755±14 | 838±1  | 917±9  | 1051±9  | 1105±1  | 1373±20 |
| 187   | 423±19 | 423±15 | 539±5  | 757±12 | 836±0  | 919±11 | 1054±6  | 1103±3  | 1366±14 |
| 189   | 402±3  | 456±17 | 536±8  | 762±8  | 840±3  | 919±11 | 1051±9  | 1113±7  | 1373±21 |
| 195   | 403±2  | 460±21 | 536±7  | 765±4  | 844±7  | 919±11 | 1051±8  | 1112±7  | 1345±7  |
| 214   | 401±4  | 422±17 | 537±6  | 757±12 | 839±2  | 916±9  | 1055±4  | 1116±10 | 1361±9  |
| 225   | 397±8  | 460±21 | 537±6  | 757±12 | 837±0  | 917±9  | 1050±10 | 1121±15 | 1373±20 |
| 226   | 396±9  | 423±16 | 536±8  | 763±6  | 837±0  | 919±11 | 1050±10 | 1120±14 | 1374±21 |
| 228   | 396±8  | 423±15 | 534±9  | 758±11 | 837±0  | 919±11 | 1053±6  | 1101±5  | 1375±23 |
| 238   | 397±8  | 423±15 | 537±7  | 761±8  | 836±0  | 919±11 | 1054±6  | 1117±11 | 1375±22 |
| 260   | 396±9  | 424±15 | 538±6  | 758±11 | 838±1  | 919±11 | 1053±7  | 1116±10 | 1377±24 |
| 268   | 398±7  | 428±11 | 536±8  | 762±7  | 837±0  | 918±10 | 1050±10 | 1118±12 | 1373±20 |
| 275   | 400±5  | 455±16 | 537±7  | 779±10 | 837±0  | 920±12 | 1051±9  | 1113±7  | 1373±21 |
| 290   | 402±3  | 453±15 | 540±4  | 761±9  | 838±1  | 905±3  | 1054±6  | 1121±16 | 1371±19 |
| 400   | 401±4  | 422±16 | 537±7  | 760±9  | 849±12 | 919±11 | 1051±9  | 1112±7  | 1374±21 |

Tabla A.7: Picos localizados de glucosa en suero.

| mg/dL | 405    | 439    | 544   | 769    | 837    | 908    | 1060    | 1106    | 1353    |
|-------|--------|--------|-------|--------|--------|--------|---------|---------|---------|
| 56    | 422±17 | 454±15 | 537±7 | 756±13 | 841±4  | 936±28 | 1052±8  | 1109±3  | 1300±53 |
| 72    | 419±14 | 453±14 | 537±7 | 753±16 | 834±3  | 936±29 | 1051±9  | 1111±5  | 1371±19 |
| 74    | 417±13 | 438±0  | 539±5 | 757±12 | 850±13 | 938±30 | 1052±7  | 1091±15 | 1378±26 |
| 80    | 424±19 | 454±15 | 536±7 | 757±12 | 839±2  | 936±29 | 1052±8  | 1108±3  | 1300±52 |
| 89    | 402±3  | 454±15 | 536±7 | 754±15 | 849±12 | 936±28 | 1052±7  | 1109±3  | 1298±54 |
| 90    | 403±2  | 453±14 | 539±4 | 760±10 | 837±0  | 909±1  | 1056±4  | 1109±3  | 1373±20 |
| 95    | 420±15 | 455±17 | 537±6 | 758±11 | 840±3  | 907±1  | 1051±9  | 1113±7  | 1330±23 |
| 99    | 397±8  | 454±15 | 535±8 | 754±15 | 839±3  | 936±28 | 1051±8  | 1111±5  | 1372±19 |
| 100   | 397±8  | 452±13 | 536±7 | 753±16 | 838±1  | 912±4  | 1052±8  | 1111±5  | 1371±19 |
| 101   | 400±5  | 453±14 | 537±6 | 753±16 | 840±3  | 937±29 | 1051±8  | 1111±5  | 1377±25 |
| 104   | 421±16 | 453±14 | 536±7 | 753±16 | 837±0  | 938±30 | 1051±9  | 1112±6  | 1373±20 |
| 106   | 422±18 | 457±19 | 537±7 | 754±15 | 844±7  | 912±4  | 1051±8  | 1113±7  | 1300±52 |
| 107   | 421±16 | 450±12 | 537±7 | 755±15 | 840±3  | 937±29 | 1052±8  | 1112±6  | 1309±43 |
| 108   | 416±11 | 464±25 | 535±9 | 754±15 | 845±8  | 938±30 | 1053±7  | 1111±5  | 1374±21 |
| 109   | 422±17 | 453±14 | 536±8 | 753±16 | 839±2  | 936±29 | 1051±9  | 1110±4  | 1372±20 |
| 112   | 422±17 | 451±12 | 535±8 | 757±12 | 840±3  | 914±6  | 1051±8  | 1112±6  | 1368±16 |
| 115   | 422±17 | 451±12 | 537±6 | 756±13 | 844±7  | 936±28 | 1050±9  | 1113±7  | 1301±51 |
| 117   | 421±16 | 459±20 | 537±6 | 754±15 | 844±7  | 937±29 | 1051±9  | 1112±6  | 1300±52 |
| 125   | 418±13 | 444±5  | 537±6 | 753±16 | 847±10 | 935±28 | 1054±6  | 1112±6  | 1372±19 |
| 126   | 427±23 | 462±23 | 540±4 | 755±14 | 846±9  | 936±29 | 1054±5  | 1115±9  | 1375±23 |
| 127   | 419±14 | 455±16 | 536±8 | 756±13 | 838±1  | 936±28 | 1052±8  | 1111±5  | 1309±44 |
| 129   | 393±12 | 434±5  | 542±2 | 769±0  | 838±1  | 900±8  | 1059±1  | 1112±6  | 1374±21 |
| 140   | 400±5  | 452±13 | 536±7 | 757±12 | 842±5  | 937±29 | 1051±9  | 1111±5  | 1376±23 |
| 145   | 424±19 | 455±17 | 537±7 | 757±12 | 841±4  | 937±29 | 1052±8  | 1110±4  | 1301±52 |
| 153   | 420±15 | 454±15 | 537±6 | 753±16 | 842±5  | 937±29 | 1051±9  | 1111±5  | 1373±20 |
| 156   | 397±8  | 453±15 | 537±6 | 755±14 | 840±3  | 937±30 | 1052±8  | 1113±7  | 1375±22 |
| 163   | 425±20 | 459±21 | 539±4 | 757±13 | 846±9  | 915±7  | 1053±7  | 1144±38 | 1303±50 |
| 165   | 419±14 | 453±14 | 538±6 | 755±14 | 837±0  | 938±30 | 1051±9  | 1111±5  | 1311±42 |
| 169   | 400±5  | 421±18 | 536±7 | 752±17 | 849±13 | 937±29 | 1051±8  | 1111±6  | 1301±52 |
| 187   | 420±15 | 451±12 | 538±6 | 754±15 | 837±0  | 937±29 | 1050±9  | 1112±6  | 1374±21 |
| 189   | 415±10 | 454±15 | 539±5 | 759±11 | 840±3  | 916±8  | 1053±7  | 1108±2  | 1374±21 |
| 195   | 419±14 | 456±17 | 537±7 | 753±16 | 842±5  | 934±26 | 1049±10 | 1110±4  | 1380±27 |
| 214   | 428±23 | 453±15 | 536±7 | 754±15 | 839±3  | 936±28 | 1052±8  | 1108±2  | 1298±54 |
| 225   | 421±16 | 461±22 | 536±7 | 752±17 | 849±13 | 937±29 | 1051±8  | 1111±6  | 1301±52 |
| 226   | 420±15 | 451±12 | 537±7 | 754±15 | 841±4  | 937±30 | 1051±9  | 1111±6  | 1301±52 |
| 228   | 416±12 | 451±12 | 537±7 | 753±16 | 840±4  | 937±29 | 1052±7  | 1113±8  | 1377±24 |
| 238   | 417±13 | 451±12 | 537±6 | 754±15 | 839±3  | 937±29 | 1051±9  | 1112±6  | 1372±20 |
| 260   | 422±17 | 455±16 | 537±6 | 756±13 | 844±7  | 936±28 | 1051±9  | 1112±6  | 1309±43 |
| 268   | 421±17 | 453±14 | 537±6 | 754±15 | 841±4  | 937±29 | 1051±9  | 1111±5  | 1301±52 |
| 275   | 425±20 | 454±15 | 537±6 | 757±12 | 839±3  | 936±29 | 1051±9  | 1111±5  | 1299±53 |
| 290   | 417±12 | 453±15 | 537±6 | 755±14 | 840±3  | 937±30 | 1052±8  | 1113±7  | 1375±22 |
| 400   | 430±25 | 455±16 | 537±6 | 757±12 | 847±10 | 919±11 | 1051±8  | 1111±6  | 1398±46 |

## PROGRAMAS

---

### B.1. Rejilla de error de Clarke

La siguiente función se obtuvo de Edgar Guevara (2021). Clarke Error Grid Analysis MATLAB Central File Exchange. Retrieved May 19, 2021.

```
function [total , percentage] = clarke(y,yp)

%SYNTAX:
%
%[total , percentage] = clarke(y,yp)
%
%INPUTS:
%y                Reference values (mg/dl)
%yp               Predicted/estimated values (mg/dl)
%
%OUTPUTS:
%total            Total points per zone:
%                 total(1) = zone A,
%                 total(2) = zone B, and so on
%percentage       Percentage of data which fell in certain region:
%                 percentage(1) = zone A,
%                 percentage(2) = zone B, and so on.
%
%EXAMPLE:         load example_data.mat
%                 [tot , per] = clarke(y,yp)
%
%Error checking
```

---

```

if nargin == 0
    error('clarke:Inputs','There_are_no_inputs.')
end
if length(yp) ~= length(y)
    error('clarke:Inputs','Vectors_y_and_yp_must_be_the_same_length.')
end
if (max(y) > 410) || (max(yp) > 410) || (min(y) < 0) || (min(yp) < 0)
    error('clarke:Inputs','Vectors_y_and_yp_are_not
    _____in_the_physiological_range_of_glucose_(<400mg/dl).')
end
%---Print figure flag-----
PRINT_FIGURE = true;
%-- Determine data length--
n = length(y);
%--- Plot Clarke's Error Grid----
h = figure;
plot(y,yp,'ko','MarkerSize',4,'MarkerFaceColor','k','MarkerEdgeColor','k');
xlabel('Concentracion_de_referencia_[mg/dl]');
ylabel('Concentracion_predicha_[mg/dl]');
title('Análisis_de_Rejilla_de_error_de_Clarke');
set(gca,'XLim',[0 400]);
set(gca,'YLim',[0 400]);
axis square
hold on
plot([0 400],[0 400],'k:')
% Theoretical 45 degrees regression line
plot([0 175/3],[70 70],'k-')
% plot([175/3 320],[70 400],'k-')
plot([175/3 400/1.2],[70 400],'k-')
% replace 320 with 400/1.2 because 100*(400 - 400/1.2)/(400/1.2) = 20% error
plot([70 70],[84 400],'k-')
plot([0 70],[180 180],'k-')
plot([70 290],[180 400],'k-')
% Corrected upper B-C boundary
% plot([70 70],[0 175/3],'k-')
plot([70 70],[0 56],'k-')

```

---



---

```

% replace 175.3 with 56 because 100*abs(56-70)/70 = 20% error
% plot([70 400],[175/3 320],'k-')
plot([70 400],[56 320],'k-')
plot([180 180],[0 70],'k-')
plot([180 400],[70 70],'k-')
plot([240 240],[70 180],'k-')
plot([240 400],[180 180],'k-')
plot([130 180],[0 70],'k-')
% Lower B-C boundary slope OK
text(30,20,'A','FontSize',12);
text(30,150,'D','FontSize',12);
text(30,380,'E','FontSize',12);
text(150,380,'C','FontSize',12);
text(160,20,'C','FontSize',12);
text(380,20,'E','FontSize',12);
text(380,120,'D','FontSize',12);
text(380,260,'B','FontSize',12);
text(280,380,'B','FontSize',12);
set(h, 'color', 'white');
% sets the color to white
% Specify window units
set(h, 'units', 'inches')
% Change figure and paper size (Fixed to 3x3 in)
set(h, 'Position', [0.1 0.1 3 3])
set(h, 'PaperPosition', [0.1 0.1 3 3])
if PRINT_FIGURE
    % Saves plot as a Enhanced MetaFile
    print(h, '-dmeta', 'Clarke_EGA');
    % Saves plot as PNG at 300 dpi
    print(h, '-dpng', 'Clarke_EGA', '-r300');
end
total = zeros(5,1); % Initializes output
%--- Statistics ----
for i=1:n,
    if (yp(i) <= 70 && y(i) <= 70) || (yp(i) <= 1.2*y(i) && yp(i) >= 0.8*y(i))
        total(1) = total(1) + 1;
    end
end

```

---

```

        % Zone A
    else
        if ( (y(i) >= 180) && (yp(i) <= 70) ) || ( (y(i) <= 70) && yp(i) >=
            total(5) = total(5) + 1;
            % Zone E
        else
            if ((y(i) >= 70 && y(i) <= 290) && (yp(i) >= y(i) + 110) ) || ((
                total(3) = total(3) + 1;
                % Zone C
            else
                if ((y(i) >= 240) && ((yp(i) >= 70) && (yp(i) <= 180))) || (
                    total(4) = total(4) + 1;
                    % Zone D
                else
                    total(2) = total(2) + 1;
                    % Zone B
                end
                % End of 4th if
            end
            % End of 3rd if
        end
        % End of 2nd if
    end
    % End of 1st if
end
% End of for loop
percentage = (total./n)*100;

```

## B.2. RReliefF-SOM

```

clear all
close all
clc

load Antebrazo_SinOutlayers.txt

```

---

```

load Antebrazo_SinOutlayersNAMES.txt
X=Antebrazo_SinOutlayers(162:673,:)' ; % 600-1600
% X1=Antebrazo_SinOutlayers' ; % 200-1800
Y=Antebrazo_SinOutlayersNAMES';

P = cvpartition(Y,'Kfold',3);
tic;
for Nei=5:40
    for ii=1:P.NumTestSets
        [ranking, w] = relieff(X1, Y, Nei);
        kk=512;
        X=X1(:,ranking(1:kk));
        trIdx=P.training(ii);
        teIdx=P.test(ii);
        X_train=X(trIdx,:)' ;
        X_test=X(teIdx,:)' ;
        Y_train=Y(trIdx,:);
        Y_test=Y(teIdx,:);

        eval(['Y_Test_Fold' num2str(ii) 'Nei' num2str(Nei)...
            '=Y_test;']);
        eval(['Y_Train_Fold' num2str(ii) 'Nei' num2str(Nei)...
            '=Y_train;']);
        %%RED SOM
        D=10;
        dimension1 = D; dimension2 = D;
        net = selforgmap([dimension1 dimension2]);
        [net, tr] = train(net, X_train);
        y = net(X_train)';
        Weight=net.iw;
        Weight1=Weight{1,1};
        [a b]=size(y);
        [c d]=size(Weight1);
        clear Weight_train
        for j=1:a
            maximum = max(max(y(j,:)));

```

---

```

        node=find((y(j,:)==maximum));
        position(j,:)=node;
        Weight_train(j,:)=Weight1(position(j),:);
    end
    % TESTING SOM
    y_Prueba1=net(X_test)';
    [h k]=size(y_Prueba1);
    clear Weight_test
    for qq=1:h
        maximum1 = max(max(y_Prueba1(qq,:)));
        node1=find((y_Prueba1(qq,:)==maximum1));
        position1(qq,:)=node1;
        Weight_test(qq,:)=Weight1(position1(qq),:);
    end
    eval(['X_Train_Fold' num2str(ii) 'Nei' ...
    num2str(Nei) '=Weight_train;']);
    eval(['X_Test_Fold' num2str(ii) 'Nei' ...
    num2str(Nei) '=Weight_test;']);
%% Neural Net Fitting
% Pureline
for o=5:20
    xP=Weight_train';
    tP = Y_train';
    trainFcnP = 'trainlm';
    % Create a Fitting Network
    hiddenLayerSizeP =o;
    netP = fitnet(hiddenLayerSizeP,trainFcnP);
    % Train the Network
    [netP,trP] = train(netP,xP,tP);
    % Test the Network
    yyP = netP(xP)';
    eP = gsubtract(tP,yyP);
    performanceP = perform(netP,tP,yyP);
    %% Testing
    Test1P=Weight_test';
    clasestestP=Y_test;

```

---

```

    t1P = clasestestP';
    y_Prueba2P=netP(Test1P)';
    eval(['Antebrazo_Purelin_N' num2str(o) 'Nei' num2str(Nei)...
    'Fold' num2str(ii) '=netP;']);
    NODOSF_P(:,o)=y_Prueba2P;
    eval(['Resultados_Pureline_N' num2str(o) 'Nei' num2str(Nei) 'Fold'...
    num2str(ii) '=NODOSF_P;']);
end
%% calculo de RMSE Pureline
[n m]=size(X_test);
for jj=5:20
    diffP=abs(NODOSF_P(:,jj)-Y_test);
    SUMAP=sumsqr(diffP);
    RMSEP(:,jj)=sqrt(SUMAP/n);
end
RMSE_nodosP=RMSEP(:,5:20);
eval(['RMSEP_Fold' num2str(ii) '=RMSE_nodosP;']);

% Tansig
for o=5:20
    xT=Weight_train';
    tT = Y_train';
    trainFcnT = 'trainlm';
    % Create a Fitting Network
    hiddenLayerSizeT =o;
    netT = fitnet(hiddenLayerSizeT,trainFcnT);
    % Activation function in layers
    netT.layers{1}.transferFcn = 'tansig';
    netT.layers{2}.transferFcn = 'tansig';
    % % Train the Network
    [netT,trT] = train(netT,xT,tT);
    % Test the Network
    yyT = netT(xT)';
    eT = gsubtract(tT,yyT);
    performanceT = perform(netT,tT,yyT);
    %% Testing

```

---

---

```

    Test1T=Weight_test';
    clasestestT=Y_test;
    t1T = clasestestT';
    y_Prueba2T=netT(Test1T)';
    eval(['Antebrazo_Tansig_N' num2str(o) 'Nei' num2str(Nei)...
    'Fold' num2str(ii) '=netT;']);
    NODOSF_T(:,o)=y_Prueba2T;
    eval(['Resultados_Tansig_N' num2str(o) 'Nei' num2str(Nei)...
    'Fold' num2str(ii) '=NODOSF_T;']);
end
    %% calculo de RMSE Tansig
for jj=5:20
    diffT=abs(NODOSF_T(:,jj)-Y_test);
    SUMAT=sumsqr(diffT);
    RMSET(:,jj)=sqrt(SUMAT/n);
end
RMSE_nodosT=RMSET(:,5:20);
eval(['RMSET_Fold' num2str(ii) '=RMSE_nodosT;']);
fprintf('Termino_el_Fold_%d_',ii)
end
disp('///fin\\\\')
%% calculo de RMSE-CV
RMSEP_F=vertcat(RMSEP_Fold1,RMSEP_Fold2,RMSEP_Fold3);
RMSET_F=vertcat(RMSET_Fold1,RMSET_Fold2,RMSET_Fold3);

eval(['RMSEP_Total_Nei' num2str(Nei) '=RMSEP_F;']);
eval(['RMSET_Total_Nei' num2str(Nei) '=RMSET_F;']);

for kk=1:16
    RMSEP_TOTAL=sum(RMSEP_F(:,kk))/3;
    RMSET_TOTAL=sum(RMSET_F(:,kk))/3;
    RMSEP_FINAL(:,kk)=RMSEP_TOTAL;
    RMSET_FINAL(:,kk)=RMSET_TOTAL;
end

STDP=std(RMSEP_F);

```

---

---

```

STDT=std(RMSET_F);

FinalSTDP=vertcat(RMSEP_FINAL,STDP);
FinalSTDt=vertcat(RMSET_FINAL,STDT);
Min_RMSEP=min(min(RMSEP_FINAL));
Min_RMSET=min(min(RMSET_FINAL));

eval([ 'STDP_Nei' num2str(Nei) '=FinalSTDP;' ]);
eval([ 'STDT_Nei' num2str(Nei) '=FinalSTDt;' ]);
eval([ 'Min_RMSEP_F_Nei' num2str(Nei) '=Min_RMSEP;' ]);
eval([ 'Min_RMSET_F_Nei' num2str(Nei) '=Min_RMSET;' ]);

Loc_RMSEP=find(RMSEP_FINAL==Min_RMSEP);
MIN_RMSE_Pu_DESVI=vertcat(FinalSTDP(:,Loc_RMSEP),(Loc_RMSEP+4))
Loc_RMSET=find(RMSET_FINAL==Min_RMSET);
MIN_RMSE-Ta_DESVI=vertcat(FinalSTDt(:,Loc_RMSET),(Loc_RMSET+4))

eval([ 'Loc_RMSEP_F' num2str(ii) 'Nei' num2str(Nei)...
' =Loc_RMSEP;' ]);
eval([ 'Loc_RMSET_F' num2str(ii) 'Nei' num2str(Nei)...
' =Min_RMSET;' ]);
fprintf('Termino_el_Neighbor_%d_',Nei)
end
toc;

```

### B.3. CFS, Wrapper-SOM

```

clear all
close all
clc

%%Probando WEKA 3 Fold cross-validation
load antebrazo_glucose.txt
load names_antebrazo.txt

```

---

```

%X=antebrazo_glucose(:,162:673); % 600-1600
X1=antebrazo_glucose; % 200-1800
y=names_antebrazo;
RankingCFS_BestF=[ 15, 152, 170, 171, 210, 218,...
219, 266, 267, 293, 302, 317, 321, 326, 381,...
413, 414, 419, 420, 512, 513, 514, 521, 535, 536,...
543, 665, 666, 694, 770, 772];
RankingCFS_Greedy=[ 15, 170, 210, 218, 219, 266,...
293, 302, 321, 326, 413, 414, 419, 513, 521,...
535, 543, 666, 694, 770, 772];
RankingWrapper_Greedy=[242, 288, 302, 467, 490, 665];
RankingWrapper_BestF=[30,54,152,160,168,242,288,302,...
309,317,318,358,381,382,417,447,458,467,490,533,573,...
630,665,731];
[mm kk]=size(RankingWrapper_BestF);
X=X1(:,RankingWrapper_BestF);
Y=y;
% pause
P = cvpartition(Y,'Kfold',3);
for ii=1:P.NumTestSets
    trIdx=P.training(ii);
    teIdx=P.test(ii);
    X_train=X(trIdx,:);
    X_test=X(teIdx,:);
    Y_train=Y(trIdx,:);
    Y_test=Y(teIdx,:);

    eval(['Y_Train_Fold' num2str(ii) 'KK_' num2str(kk)...
    '=Y_train;']);
    eval(['Y_Test_Fold' num2str(ii) 'KK_' num2str(kk)...
    '=Y_test;']);
    eval(['X_Train_Fold' num2str(ii) 'KK_' num2str(kk)...
    '=X_train;']);
    eval(['X_Test_Fold' num2str(ii) 'KK_' num2str(kk)...
    '=X_test;']);
    %%RED SOM

```

---



---

```

D=10;
dimension1 = D; dimension2 = D;
net = selforgmap([ dimension1 dimension2 ]);
[net, tr] = train(net, X_train);
y = net(X_train)';
Weight=net.iw;
Weight1=Weight{1,1};
[a b]=size(y);
[c d]=size(Weight1);
clear Weight_train
for j=1:a
    maximum = max(max(y(j, :)));
    node=find((y(j, :)==maximum));
    position(j, :)=node;
    Weight_train(j, :)=Weight1(position(j), :);
end
% TESTING SOM
y_Prueba1=net(X_test)';
[h k]=size(y_Prueba1);
clear Weight_test
for uu=1:h
    maximum1 = max(max(y_Prueba1(uu, :)));
    node1=find((y_Prueba1(uu, :)==maximum1));
    position1(uu, :)=node1;
    Weight_test(uu, :)=Weight1(position1(uu), :);
end
eval(['XWeight_Train_Fold' num2str(ii) 'KK_' ...
num2str(kk) '=Weight_train;']);
eval(['XWeight_Test_Fold' num2str(ii) 'KK_' ...
num2str(kk) '=Weight_test;']);
%% Neural Net Fitting
%%Pureline
for o=5:20
    xP=Weight_train';
    tP = Y_train';
    trainFcnP = 'trainlm';

```

---

```

% Create a Fitting Network
hiddenLayerSizeP =0;
netP = fitnet(hiddenLayerSizeP ,trainFcnP);
% % Train the Network
[netP ,trP] = train(netP ,xP ,tP );
% Test the Network
yyP = netP(xP)';
eP = gsubtract(tP ,yyP);
performanceP = perform(netP ,tP ,yyP);
%%Testing
Test1P=Weight_test';
clasestestP=Y_test;
t1P = clasestestP';
y_Prueba2P=netP(Test1P)';
eval(['Antebrazo_Fold' num2str(ii) 'SOM_Purelin_N' num2str(o)...
'KK_' num2str(kk) '=netP;']);
NODOSF_P(:,o)=y_Prueba2P;
eval(['Resultados_Pureline_Fold' num2str(ii) '_N' num2str(o) ...
'KK_' num2str(kk) '=NODOSF_P;']);
dlmwrite(['Resultados_Pureline_kk' num2str(kk) '_Fold'...
num2str(ii) '.txt'],NODOSF_P,'delimiter ','\t');
end
%%calculo de RMSE Pureline
[n m]=size(Weight_test);
for jj=5:20
    diffP=abs(NODOSF_P(:,jj)-Y_test);
    SUMAP=sumsqr(diffP);
    RMSEP(:,jj)=sqrt(SUMAP/n);
end
matrizP=floor(linspace(5,20,16));
RMSE_nodosP=RMSEP(:,5:20);
eval(['RMSEP_Fold' num2str(ii) '=RMSE_nodosP;']);

%%Tansig
for o=5:20
    xT=Weight_train';

```

---

```

tT = Y_train';
trainFcnT = 'trainlm';
% Create a Fitting Network
hiddenLayerSizeT = 0;
netT = fitnet(hiddenLayerSizeT, trainFcnT);
% Activation function in layers
netT.layers{1}.transferFcn = 'tansig';
netT.layers{2}.transferFcn = 'tansig';
% % Train the Network
[netT, trT] = train(netT, xT, tT);
% Test the Network
yyT = netT(xT)';
eT = gsubtract(tT, yyT);
performanceT = perform(netT, tT, yyT);
%%Testing
Test1T = Weight_test';
claseTestT = Y_test;
t1T = claseTestT';
y_Prueba2T = netT(Test1T)';
eval(['Antebrazo_Fold' num2str(ii) 'SOM_Tansig_N' ...
num2str(o) 'KK_' num2str(kk) '=netT;']);
NODOSF_T(:, o) = y_Prueba2T;
eval(['Resultados_Tansig_Fold' num2str(ii) '_N' ...
num2str(o) 'KK_' num2str(kk) '=NODOSF_T;']);
dlmwrite(['Resultados_Tansig_kk' num2str(kk) ...
'_Fold' num2str(ii) '.txt'], NODOSF_T, 'delimiter', '\t');
end
%% calculo de RMSE Tansig
for jj = 5:20
    diffT = abs(NODOSF_T(:, jj) - Y_test);
    SUMAT = sumsq(diffT);
    RMSET(:, jj) = sqrt(SUMAT/n);
end
matriz = floor(linspace(5, 20, 16));
RMSE_nodosT = RMSET(:, 5:20);
eval(['RMSET_Fold' num2str(ii) '=RMSE_nodosT;']);

```

---

---

```

    fprintf('Termino el Fold %d ',ii)
end
RMSEP_F=vertcat(RMSEP_Fold1,RMSEP_Fold2,RMSEP_Fold3);
RMSET_F=vertcat(RMSET_Fold1,RMSET_Fold2,RMSET_Fold3);
for kn=1:16
    RMSEP_TOTAL=sum(RMSEP_F(:,kn))/3;
    RMSET_TOTAL=sum(RMSET_F(:,kn))/3;
    RMSEP_FINAL(:,kn)=RMSEP_TOTAL;
    RMSET_FINAL(:,kn)=RMSET_TOTAL;
end
STDP=std(RMSEP_F);
STDT=std(RMSET_F);
FinalSTDP=vertcat(RMSEP_FINAL,STDP);
FinalSTDT=vertcat(RMSET_FINAL,STDT);
Min_RMSEP=min(min(RMSEP_FINAL));
Min_RMSET=min(min(RMSET_FINAL));

Loc_RMSEP=find(RMSEP_FINAL==Min_RMSEP);
MIN_RMSE_Pu_DESVI=vertcat(FinalSTDP(:,Loc_RMSEP),(Loc_RMSEP+4))
Loc_RMSET=find(RMSET_FINAL==Min_RMSET);
MIN_RMSE-Ta_DESVI=vertcat(FinalSTDT(:,Loc_RMSET),(Loc_RMSET+4))

```

## B.4. PCA-SOM

```

clear all
close all
clc

load antebrazo_glucose.txt
load names_antebrazo.txt
X1=antebrazo_glucose(:,162:673); %600-1600
% X1=antebrazo_glucose; %200-1800
Y=names_antebrazo;
%%PCA
STR=Y;

```

---

```

CHR=string (STR);
names2=char (CHR);
Es=200;
[coeff ,score ,latent ,~,explained]=pca (X1);
suma=sum(explained (1:100))
X=score (:,1:100);
[mmz kk]=size (X)
c=score (:,1);
d=score (:,2);
f=score (:,3);
% [ranking , w] = relieff( score , Y, 26);
% kk=100;
%X=score (:,ranking (1:kk));
% suma=sum(explained (ranking (1:kk)))

P = cvpartition (Y, 'Kfold' ,3);
for ii =1:P.NumTestSets
    trIdx=P.training (ii);
    teIdx=P.test (ii);
    X_train=X(trIdx ,:)' ;
    X_test=X(teIdx ,:)' ;
    Y_train=Y(trIdx ,:);
    Y_test=Y(teIdx ,:);

    eval ([ 'Y_Test_Fold' num2str(ii) '=Y_test;' ]);
    eval ([ 'Y_Train_Fold' num2str(ii) '=Y_train;' ]);
    % RED SOM
    D=10;
    dimension1 = D; dimension2 = D;
    net = selforgmap ([dimension1 dimension2]);
    [net ,tr] = train(net ,X_train);
    y = net(X_train)';
    Weight=net.iw;
    Weight1=Weight {1 ,1};
    [a b]=size (y);
    [c d]=size (Weight1);

```

---

```

clear Weight_train
for j=1:a
    maximum = max(max(y(j,:)));
    node=find((y(j,:)==maximum));
    position(j,:)=node;
    Weight_train(j,:)=Weight1(position(j,:),:);
end
% TESTING SOM
y_Prueba1=net(X_test)';
[h k]=size(y_Prueba1);
clear Weight_test
for qq=1:h
    maximum1 = max(max(y_Prueba1(qq,:)));
    node1=find((y_Prueba1(qq,:)==maximum1));
    position1(qq,:)=node1;
    Weight_test(qq,:)=Weight1(position1(qq,:),:);
end
eval(['X_Train_Fold' num2str(ii) '=Weight_train;']);
eval(['X_Test_Fold' num2str(ii) '=Weight_test;']);
%% Neural Net Fitting
% Pureline
for o=5:20
    xP=Weight_train';
    tP = Y_train';
    trainFcnP = 'trainlm';
    % Create a Fitting Network
    hiddenLayerSizeP =o;
    netP = fitnet(hiddenLayerSizeP,trainFcnP);
    % Train the Network
    [netP,trP] = train(netP,xP,tP);
    % Test the Network
    yyP = netP(xP)';
    eP = gsubtract(tP,yyP);
    performanceP = perform(netP,tP,yyP);
    %%Testing
    Test1P=Weight_test';

```

---

```

        clasestestP=Y_test;
        t1P = clasestestP';
        y_Prueba2P=netP(Test1P)';
        eval(['Antebrazo_Purelin_N' num2str(o) ...
        'Fold' num2str(ii) '=netP;']);
        NODOSF_P(:,o)=y_Prueba2P;
        eval(['Resultados_Pureline_N' num2str(o) ...
        'Fold' num2str(ii) '=NODOSF_P;']);
        dlmwrite(['ResultadosReliefF_Pureline_PC' num2str(kk) ...
        '_Fold' num2str(ii) '.txt'],NODOSF_P,'delimiter ','\t');
    end
    %%calculo de RMSE Pureline
    [n m]=size(X_test);
    for jj=5:20
        diffP=abs(NODOSF_P(:,jj)-Y_test);
        SUMAP=sumsqr(diffP);
        RMSEP(:,jj)=sqrt(SUMAP/n);
    end
    matrizP=floor(linspace(5,20,16));
    RMSE_nodosP=RMSEP(:,5:20);
    eval(['RMSEP_Fold' num2str(ii) '=RMSE_nodosP;']);
    % Tansig
    for o=5:20
        xT=Weight_train';
        tT = Y_train';
        trainFcnT = 'trainlm';
        % Create a Fitting Network
        hiddenLayerSizeT =o;
        netT = fitnet(hiddenLayerSizeT,trainFcnT);
        % Activation function in layers
        netT.layers{1}.transferFcn = 'tansig';
        netT.layers{2}.transferFcn = 'tansig';
        % % Train the Network
        [netT,trT] = train(netT,xT,tT);
        % Test the Network
        yyT = netT(xT)';
    end

```

---

```

        eT = gsubtract(tT,yyT);
        performanceT = perform(netT,tT,yyT);
        %%Testing
        Test1T=Weight_test';
        clasestestT=Y_test;
        t1T = clasestestT';
        y_Prueba2T=netT(Test1T)';
        eval(['Antebrazo_Tansig_N' num2str(o) 'Fold'...
        num2str(ii) '=netT;']);
        NODOSF_T(:,o)=y_Prueba2T;
        eval(['Resultados_Tansig_N' num2str(o) 'Fold'...
        num2str(ii) '=NODOSF_T;']);
        dlmwrite(['ResultadosReliefF_Tansig_PC'... num2str(kk) '_Fold' n
end
        %%calculo de RMSE Tansig
        for jj=5:20
            diffT=abs(NODOSF_T(:,jj)-Y_test);
            SUMAT=sumsqr(diffT);
            RMSET(:,jj)=sqrt(SUMAT/n);
        end
        matriz=floor(linspace(5,20,16));
        RMSE_nodosT=RMSET(:,5:20);
        eval(['RMSET_Fold' num2str(ii) '=RMSE_nodosT;']);

end
disp('///fin\\\\')
RMSEP_F=vertcat(RMSEP_Fold1,RMSEP_Fold2,RMSEP_Fold3);
RMSET_F=vertcat(RMSET_Fold1,RMSET_Fold2,RMSET_Fold3);

for kk=1:16
    RMSEP_TOTAL=sum(RMSEP_F(:,kk))/3;
    RMSET_TOTAL=sum(RMSET_F(:,kk))/3;

    RMSEP_FINAL(:,kk)=RMSEP_TOTAL;
    RMSET_FINAL(:,kk)=RMSET_TOTAL;
end

```

---



```

STDP=std (RMSEP_F);
STDT=std (RMSET_F);
FinalSTDP=vertcat (RMSEP_FINAL,STDP);
FinalSTDT=vertcat (RMSET_FINAL,STDT);
Min_RMSEP=min (min (RMSEP_FINAL));
Min_RMSET=min (min (RMSET_FINAL));

eval ([ 'STDP_Nei' ' '=FinalSTDP; ']);
eval ([ 'STDT_Nei' ' '=FinalSTDT; ']);
eval ([ 'Min_RMSEP_F_Nei' ' '=Min_RMSEP; ']);
eval ([ 'Min_RMSET_F_Nei' ' '=Min_RMSET; ']);

Loc_RMSEP=find (RMSEP_FINAL==Min_RMSEP);
MIN_RMSE_Pu_DESVI=vertcat (FinalSTDP (: , Loc_RMSEP) ,(Loc_RMSEP+4))
Loc_RMSET=find (RMSET_FINAL==Min_RMSET);
MIN_RMSE-Ta_DESVI=vertcat (FinalSTDT (: , Loc_RMSET) ,(Loc_RMSET+4))

eval ([ 'Loc_RMSEP_F' num2str (ii) ' '=Loc_RMSEP; ']);
eval ([ 'Loc_RMSET_F' num2str (ii) ' '=Min_RMSET; ']);

```

## B.5. SVM-LR

```

clear all
clc
close all

%%Probando SVM y LR
load X_Train_Fold1Nei9.mat
load X_Train_Fold2Nei9.mat
load X_Train_Fold3Nei9.mat
load Y_Train_Fold1Nei9.mat
load Y_Train_Fold2Nei9.mat
load Y_Train_Fold3Nei9.mat

```

```
load X_Test_Fold1Neig.mat
load X_Test_Fold2Neig.mat
load X_Test_Fold3Neig.mat
load Y_Test_Fold1Neig.mat
load Y_Test_Fold2Neig.mat
load Y_Test_Fold3Neig.mat

Fold1_Train=horzcat(X_Train_Fold1Neig, Y_Train_Fold1Neig);
Fold2_Train=horzcat(X_Train_Fold2Neig, Y_Train_Fold2Neig);
Fold3_Train=horzcat(X_Train_Fold3Neig, Y_Train_Fold3Neig);
pause
load trainedModel_SVMF1.mat
load trainedModel_SVMF2.mat
load trainedModel_SVMF3.mat
%%          SVM
%%Fold 1
yfit1 = trainedModel_SVMF1.predictFcn(X_Test_Fold1Neig);
[m n]=size(Y_Test_Fold1Neig);
diffT=abs(yfit1-Y_Test_Fold1Neig);
SUMAT=sumsqr(diffT);
RMSE1=sqrt(SUMAT/m);
%%Fold 2
yfit2 = trainedModel_SVMF2.predictFcn(X_Test_Fold2Neig);
[m n]=size(Y_Test_Fold2Neig);
diffT=abs(yfit2-Y_Test_Fold2Neig);
SUMAT=sumsqr(diffT);
RMSE2=sqrt(SUMAT/m);
%%Fold 3
yfit3 = trainedModel_SVMF3.predictFcn(X_Test_Fold3Neig);
[m n]=size(Y_Test_Fold3Neig);
diffT=abs(yfit3-Y_Test_Fold3Neig);
SUMAT=sumsqr(diffT);
RMSE3=sqrt(SUMAT/m);

%%Total
RMSE_F=vertcat(RMSE1,RMSE2,RMSE3);
```

```

RMSE_TOTAL=sum(RMSE_F)/3
STD=std(RMSE_F)

%%                               LR
load trainedModel_LRF1.mat
load trainedModel_LRF2.mat
load trainedModel_LRF3.mat
%%Fold 1
yfit1 = trainedModel_LRF1.predictFcn(X_Test_Fold1Neig);
[m n]=size(Y_Test_Fold1Neig);
diffT=abs(yfit1-Y_Test_Fold1Neig);
SUMAT=sumsqr(diffT);
RMSE1=sqrt(SUMAT/m);
%%Fold 2
yfit2 = trainedModel_LRF2.predictFcn(X_Test_Fold2Neig);
[m n]=size(Y_Test_Fold2Neig);
diffT=abs(yfit2-Y_Test_Fold2Neig);
SUMAT=sumsqr(diffT);
RMSE2=sqrt(SUMAT/m);
%%Fold 3
yfit3 = trainedModel_LRF3.predictFcn(X_Test_Fold3Neig);
[m n]=size(Y_Test_Fold3Neig);
diffT=abs(yfit3-Y_Test_Fold3Neig);
SUMAT=sumsqr(diffT);
RMSE3=sqrt(SUMAT/m);

%%Total
RMSE_F=vertcat(RMSE1,RMSE2,RMSE3);
RMSE_TOTAL=sum(RMSE_F)/3
STD=std(RMSE_F)

```

## B.6. Clasificación

```
clear all
```

---

```

load Antebraco_Clases.txt
load Clases.txt
load NumClases.txt
load Antebraco_SinOutlayers.txt
% x1 = Antebraco_Clases';
% x=Antebraco_SinOutlayers;
x=Antebraco_SinOutlayers;
t = Clases';

for ii=1:10
trainFcn = 'trainscg';
hiddenLayerSize = 10;
net = patternnet(hiddenLayerSize, trainFcn);
net.divideParam.trainRatio = 70/100;
net.divideParam.valRatio = 15/100;
net.divideParam.testRatio = 15/100;
% Train the Network
[net,tr] = train(net,x,t);
eval(['Rep' num2str(ii) '=net;']);
% Test the Network
y = net(x);
e = gsubtract(t,y);
performance = perform(net,t,y);
tind = vec2ind(t);
yind = vec2ind(y);
percentErrors = (sum(tind ~= yind)/numel(tind)*100);
percentAccu = (sum(tind == yind)/numel(tind)*100);
eval(['Accuracy' num2str(ii) '=percentAccu;']);
% Plots
figure, plotroc(t,y)
ii
end
Total=vertcat(Accuracy1,Accuracy2,Accuracy3,Accuracy4...
,Accuracy5,Accuracy6,Accuracy7,Accuracy8,Accuracy9,Accuracy10);
AccuProm=sum(Total)/10
STDP=std(Total)

```

---

## CONSENTIMIENTO INFORMADO

A continuación, se muestra un ejemplo del consentimiento informado que cada voluntario firmó antes de la realización de las mediciones mediante espectroscopia Raman, por confidencialidad de los voluntarios no se presentaron los consentimientos informados firmados.



### CONSENTIMIENTO INFORMADO PARA LA REALIZACIÓN DE ESTUDIOS NO INVASIVOS DE HEMOGLOBINA GLICOSILADA Y GLUCOSA

El estudio comprende una medición en la muñeca, antebrazo y dedo índice derecho, con un láser de baja potencia que no tiene ningún efecto secundario ni riesgo.

El (la) que suscribe, ACEPTA VOLUNTARIAMENTE realizarse el estudio no invasivo de hemoglobina glicosilada y glucosa, en el pleno uso de su facultad, y de manera voluntaria.

Manifiesto que he recibido información verbal acerca del uso de la técnica implementada y me han reiterado el compromiso de que toda la información por mi proporcionada se mantendrá confidencialmente, por lo que no tengo inconveniente en que se tomen estas mediciones.

Diabético: \_\_\_\_\_  
Edad: \_\_\_\_\_  
Teléfono: \_\_\_\_\_

No diabético: \_\_\_\_\_  
Ocupación: \_\_\_\_\_

ATENTAMENTE

\_\_\_\_\_  
Nombre y firma

\_\_\_\_\_  
Fecha  
(Tuxtla, Gutiérrez, Chiapas)

**Figura C.1:** Consentimiento informado para mediciones no invasivas.

---

## ALGUNOS MÉTODOS MÉDICOS

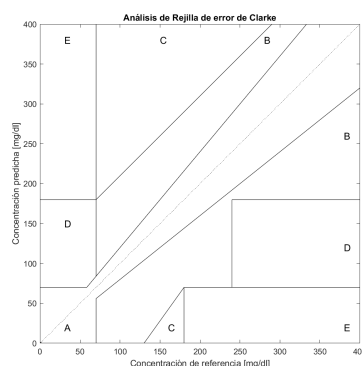
---

### D.1. Rejilla de Error de Clarke

La rejilla de Error de Clarke o también conocida como “Error Grid Analysis”, fue desarrollada en 1987 por William L. Clarke, Daniel Cox, Linda A. Gonder-Frederick, William Carter y Stephen L. Pohl, está propuesta describe la precisión clínica de los sistemas de automonitoreo de glucosa en sangre (SMBG, por sus siglas en inglés), en todo el rango de valores de glucosa en sangre.[155].

Lo que considera esta rejilla de error es: el valor absoluto del valor de glucosa generado por el sistema, el valor absoluto del valor de glucosa en sangre de referencia, la diferencia relativa entre estos dos valores, y la importancia clínica de esta diferencia, Esta rejilla define el eje  $x$  como la glucosa en sangre de referencia y el eje  $y$  como el valor generado por el sistema de propuesto. La diagonal representa una concordancia perfecta entre los dos, con puntos de datos por encima y por debajo de la diagonal que representan sobreestimaciones y subestimaciones, respectivamente [13].

En la figura D.1 se puede observar la rejilla de análisis de error, donde: los valores que se encuentran en las zonas A y B representan resultados precisos o aceptables; la zona C podría conducir a un mal resultado; la zona D representa una falla peligrosa para detectar y/o tratar y la zona E significa “tratamiento erróneo” [155, 13].



**Figura D.1:** Rejilla de análisis de error para la evaluación de las implicaciones clínicas de los valores de glucosa[13].

## D.2. Método de glucosa oxidasa

El método de glucosa oxidasa, se utiliza para la determinación cuantitativa la glucosa en plasma, suero sanguíneo u orina. Para esta prueba se requieren algunos reactivos a diferentes concentraciones como, por ejemplo: solución conteniendo glucosa oxidasa (GOD), peroxidasa (POD), 4-aminofenazona (4-AF), buffer fosfatos pH 7,0 y 4-hidroxibenzoato [141]. La obtención del suero sanguíneo se realiza mediante la extracción de sangre, la cual se guarda en un tubo con anticoagulante (EDTA/fluoruro), generalmente con tapa morado. Es importante centrifugar la sangre dentro de las 2 horas de la extracción puesto que la glucosa solo es estable 4 horas a temperatura ambiente o 24 horas refrigerada [141]. Se necesita una cantidad de muestra de 10 µl de suero, posteriormente se le agregan los reactivos antes mencionados, la glucosa en el suero generará un cambio de coloración la cual se podrá medir con un espectrofotómetro a 505 nm, esta reacción es estable durante 30 minutos. El calculo que se realiza para determinar la cantidad de glucosa se muestra en la ecuación D.1.

$$\frac{\text{Absorbancia de la muestra}}{\text{Absorbancia de referencia}} \times 100 = \text{mg/dL de glucosa} \quad (\text{D.1})$$

La sensibilidad aproximada de los espectrofotómetros a 505 nm es de 0.3 mg/dL [168].

### D.3. Método de afinidad por boronato

Existen sistemas de determinación rápida de HbA<sub>1c</sub> que proporciona resultados inmediatos. En uno de ellos se utiliza la afinidad por boronato, en este se necesita un volumen de muestra que oscila entre 1.5 a 10 µl aproximadamente y el resultado se obtiene entre 3 a 10 minutos. [169].

En este proceso se diluye la muestra automáticamente y se mezcla con una solución que lisa los eritrocitos, liberando hemoglobina. Esta mezcla de muestra se transfiere a un conjugado de ácido borónico marcado en azul, que se une a los cisdiolos de la hemoglobina glicada. Esta mezcla de reacción pasa a través de una membrana de filtro que recoge la hemoglobina unida y no unida al conjugado (hemoglobina glucosilada y no glucosilada). El exceso de conjugado se elimina con un reactivo de lavado. El dispositivo cuenta con un medidor de reflectancia que mide por separado las intensidades de color azul (hemoglobina glucosilada) y rojo (hemoglobina total). La relación de las 2 intensidades de color es proporcional al porcentaje de HbA<sub>1c</sub> en la muestra [170]



## HIPER-PARÁMETROS DE LOS ALGORITMOS DE ENTRENAMIENTO DE LAS REDES NEURONALES ARTIFICIALES

---

“trainlm” es una función de entrenamiento de red que actualiza el peso y los estados de sesgo de acuerdo con la optimización de Levenberg-Marquardt. trainlm es a menudo el algoritmo de retropropagación más rápido en la caja de herramientas y es muy recomendable como un algoritmo supervisado de primera elección, aunque requiere más memoria que otros algoritmos. Para más información visite: <https://la.mathworks.com/help/deeplearning/ug/choose-a-multilayer-neural-network-training-function.html>

### E.1. Levenberg-Marquardt

Parámetros usados en regresión

### E.2. Scaled Conjugate Gradient

“traincg” es una función de entrenamiento de red que actualiza los valores de ponderación y sesgo de acuerdo con el método de gradiente conjugado escalado. para más información visite: <https://la.mathworks.com/help/deeplearning/ref/trainscg.html>

**Tabla E.1:** Parámetros algoritmo de entrenamiento Levenberg-Marquardt usado para los experimentos reportados.

| Levenberg-Marquardt ( trainlm) |   |
|--------------------------------|---|
| epochs                         | 1000 Número máximo de épocas para entrenar              |
| goal                           | 0 Objetivo de rendimiento                               |
| max_fail                       | 6 Fallos máximos de validación                          |
| min_grad                       | 1e-7 Gradiente de rendimiento mínimo                    |
| mu                             | 0.001 Mu inicial  |
| mu_dec                         | 0.1 Factor de disminución de mu                         |
| mu_inc                         | 10 Factor de aumento de mu                              |
| mu_max                         | 1e10 Mu máximo  |
| show                           | 25 Épocas entre exhibiciones                            |
| showCommandLine                | "false" Generar salida de línea de comandos             |
| showWindow                     | "true" Mostrar interfaz gráfica de usuario de formación |
| time                           | "inf" Tiempo máximo para entrenar en segundos           |

**Tabla E.2:** Hiper-parámetros de entrenamiento usados en regresión

| Parámetros                                   | Ajustes             |
|--|---------------------|
| Número de capas escondidas                   | 1                   |
| Número de nodos en la capa escondida         | 5:1:20              |
| Función de activación para la capa escondida | tansig / pureline   |
| Función de activación para la capa de salida | pureline            |
| Algoritmo de entrenamiento                   | Levenberg-Marquardt |
| Función de costo                             | MSE                 |

**Tabla E.3:** Parámetros algoritmo de entrenamiento gradiente conjugado escalado usados para los experimentos reportados.

| Scaled conjugate gradient (trainscg) |  |
|--------------------------------------|--|
| epochs                               | 1000 Número máximo de épocas para entrenar                                     |
| show                                 | 25 Épocas entre exhibiciones   |
| showCommandLine                      | false Generar salida de línea de comandos                                      |
| showWindow                           | true Mostrar interfaz gráfica de usuario de formación                          |
| goal                                 | 0 Objetivo de rendimiento  |
| time                                 | inf Tiempo máximo para entrenar en segundos                                    |
| min_grad                             | 1e-6 Gradiente de rendimiento mínimo   |
| max_fail                             | 6 Fallos máximos de validación   |
| sigma                                | 5.0e-5 Determina el cambio de peso para la aproximación de la segunda derivada |
| lambda                               | 5.0e-7 Parámetro para regular la indefinición de la arpillera matriz Hessiana. |

**Tabla E.4:** Hiper-parámetros de entrenamiento usados en clasificación.

| Parámetros                                   | Ajustes                   |
|--|---------------------------|
| Número de capas escondidas                   | 1                         |
| Número de nodos en la capa escondida         | 5:1:20                    |
| Función de activación para la capa escondida | tansig / pureline         |
| Función de activación para la capa de salida | pureline                  |
| Algoritmo de entrenamiento                   | Scaled conjugate gradient |
| Función de costo                             | Entropía cruzada          |