

## **Estimación de valores de hemoglobina a partir del análisis de espectros FTIR de muestras de saliva de personas diabéticas**

Miguel Sánchez Brito<sup>1</sup>, Mónica M. Mata Miranda<sup>2</sup>,  
Gustavo J. Vázquez Zapién<sup>2</sup>

<sup>1</sup> TecNM/Instituto Tecnológico de Aguascalientes,  
México

<sup>2</sup> Escuela Militar de Medicina,  
Centro Militar de Ciencias de la Salud,  
Secretaría de la Defensa Nacional,  
México

{miguel\_sanchezbrito, gus1202}@hotmail.com,  
mmcmariabel@gmail.com

**Resumen.** La mala alimentación en adición con factores como el sedentarismo ha posicionado a la diabetes como una de las cuatro principales enfermedades no transmisibles a nivel mundial de acuerdo con la Organización Mundial de la Salud (OMS), siendo la diabetes tipo 2 la variante más frecuente de esta enfermedad [1]. La Asociación Americana de Diabetes (ADA) señala, esta enfermedad no es mortal, sin embargo, para evitar complicaciones que deriven en tal conclusión además de tener una buena calidad de vida es crucial un monitoreo continuo de los niveles de glucosa en sangre de los pacientes. De acuerdo con la ADA, una de las metodologías más confiables para monitorear los niveles de glucosa es la llamada prueba de hemoglobina glicosilada (A1C), la cual analiza el nivel de glicosilación de la hemoglobina (proteína exclusiva del torrente sanguíneo). Las molestias derivadas de un prolongado periodo de ayuno, así como la necesidad de repetir la prueba en más de una ocasión han permitido realizar investigaciones con la finalidad de proponer metodologías alternativas para asistir en la tarea de monitoreo de la glucosa en sangre [2]. En el presente trabajo evaluamos la posibilidad de emplear espectros obtenidos mediante espectroscopia infrarroja por transformada de Fourier (FTIR) de muestras de saliva para estimar los niveles de hemoglobina en pacientes previamente diagnosticados con diabetes tipo 2, para ello empleamos diversas técnicas comúnmente empleadas en el área de aprendizaje máquina.

**Palabras clave:** Aprendizaje máquina, espectroscopia FTIR, saliva, diabetes, hemoglobina.

### **Hemoglobin Values Estimation from the FTIR Spectra Analysis of Saliva Samples of Diabetic People**

**Abstract.** Poor diet in addition to factors such as sedentary lifestyle has positioned diabetes as one of the four main non-communicable diseases worldwide, according to the World Health Organization (WHO), with type 2

diabetes being the most frequent variant of this disease [1]. The American Diabetes Association (ADA) points out, this disease is not fatal, however, to avoid complications that lead to such a conclusion, in addition to having a good quality of life, continuous monitoring of patients' blood glucose levels is crucial. According to the ADA, one of the most reliable methodologies for monitoring glucose levels is the called glycosylated hemoglobin test (A1C), which analyzes the level of glycosylation of hemoglobin (a protein unique to the bloodstream). The discomfort derived from a prolonged period of fasting, as well as the need to repeat the test on more than one occasion, have made it possible to carry out research with the aim of proposing alternative methodologies to assist in the task of monitoring blood glucose [2]. In the present work we evaluate the possibility of using Fourier transform infrared (FTIR) spectra of saliva samples to estimate hemoglobin levels in patients previously diagnosed with type 2 diabetes, for which we use various techniques commonly used in the area of machine learning.

**Keywords:** Machine learning, FTIR spectroscopy, saliva, diabetes, hemoglobin.

## 1. Introducción

La diabetes se ha colocado como una de las principales enfermedades no transmisibles a nivel mundial de acuerdo con lo señalado por la OMS afectando alrededor de 422 millones de personas en todo el mundo [1].

Una persona con diabetes tiene de dos a tres veces más probabilidades de sufrir afectaciones cardíacas, en mujeres diabéticas embarazadas un mal control de la diabetes puede provocar muerte fetal y otras complicaciones, además, la diabetes es causante de fallas renales, amputaciones de extremidades, pérdidas de visión y afectaciones en las terminales nerviosas de los pacientes.

La diabetes es una enfermedad crónica, compleja, que requiere el cuidado médico continuo, con estrategias multifactoriales en la reducción del riesgo y control glicémico, de acuerdo con la Asociación Americana de Diabetes (ADA) las cuatro metodologías para llevar a cabo un correcto control glicémico son, la prueba A1C, *Fasting Plasma Glucose* (FPG), *Oral Glucose Tolerance Test* (OGTT), y *Random Plasma Glucose Test* (RPGT) [2]. A pesar de que ninguna de las cuatro metodologías mencionadas es considerada como *gold standard*, la prueba A1C es empleada con mayor frecuencia debido al tiempo de vida de la hemoglobina en el cuerpo humano, entre 2 y 3 meses aproximadamente, lo que permite conocer un historial de ese lapso de tiempo del comportamiento de los valores de glucosa de los pacientes a diferencia de las otras pruebas que reflejan el nivel de glucosa en el momento de la toma de la muestra. De acuerdo a lo publicado por [3], el tiempo estimado para la obtención de resultados de la prueba A1C va de entre 6 y 9 minutos y debido a la necesidad de mezclar la muestra con reactivos en conjunto con diversos pre procesamientos particulares el costo estimado de la misma es de entre 7 y 9 dólares además de requerir personal capacitado para realizar los procedimientos necesarios.

Si bien la hemoglobina es una proteína exclusiva del torrente sanguíneo, es posible encontrar otras proteínas que tienen la capacidad de vincularse con moléculas de glucosa (glicosilación) en distintos fluidos corporales como la saliva [4, 7].

La espectroscopia FTIR es una técnica que permite conocer la composición molecular de una muestra a partir de las vibraciones los enlaces de los elementos

químicos que la conforman; la muestra es sometida a radiación con frecuencias (Hertz) electromagnéticas pertenecientes a la región del espectro infrarrojo [8, 9], específicamente para el análisis de muestras biológicas se emplean frecuencias de la región media del infrarrojo (de 11.9 a 119.9 THz aproximadamente) [10, 11, 12].

La interacción de la muestra con las distintas frecuencias provoca las vibraciones de los enlaces que la conforman, cada enlace vibra con una frecuencia específica en función de los elementos que vincula y de su estructura (enlace sencillo, doble o triple); estas vibraciones se registran en un vector denominado *espectro FTIR*. El espectro FTIR de una muestra biológica ha sido ya segmentado en distintas regiones atribuidas a los principales grupos macromoleculares que la conforman: lípidos, proteínas y carbohidratos y ácidos nucleicos [13, 14].

A partir del análisis de los espectros FTIR mediante técnicas comúnmente empleadas en el área de aprendizaje máquina, se propone en el presente trabajo una metodología no invasiva para su asociación con valores de hemoglobina registrados por 258 pacientes. Los resultados obtenidos sugieren la viabilidad de emplear espectroscopia FTIR con muestras de saliva para estimar los niveles de hemoglobina de un paciente, al no requerir de reactivos ni pre procesamientos complejos la metodología propuesta podría ser una opción más ágil y accesible que la forma tradicional de realizar el estudio mediante muestras de sangre.

## 2. Métodos

En el periodo de 2019-2020 se recolectaron 258 muestras de saliva de pacientes previamente diagnosticados con diabetes tipo 2 en la Unidad de Especialidades Médicas (UEM) de la Secretaría de la Defensa Nacional (SEDENA). Los criterios de inclusión fueron pacientes de más de 18 años de edad que tuvieran un periodo de ayuno de cuando menos 8 horas.

Los criterios de exclusión contemplados fueron pacientes que se hubiera cepillado o enjuagado la cavidad oral con enjuague bucal antes de la toma de la muestra y pacientes que tuvieran algún tratamiento de ortodoncia u otro tratamiento oral.

Las muestras de saliva se procesaron con el espectrómetro Jasco FTIR-6600 empleando el accesorio *attenuated total reflectance* (ATR), con una resolución de 4  $\text{cm}^{-1}$  y 120 escaneos como sugiere [8] y posteriormente normalizados mediante la técnica *standard normal variate* (SNV) (1) siguiendo lo señalado por [15]. Empleando la configuración mencionada se obtuvieron espectros constituidos por 3736 variables o números de onda que representan las frecuencias con las que fue analizada la muestra:

$$Z = \frac{x-\mu}{\sigma}. \quad (1)$$

En (1),  $x$  es un valor real obtenido mediante espectroscopia FTIR,  $\mu$  es la media del espectro que se analiza y  $\sigma$  la desviación estándar de la señal obtenida (espectro).

La complejidad del uso de espectros FTIR en el diagnóstico clínico radica en la naturaleza de las muestras orgánicas y en la cantidad de componentes de las mismas, al compartir gran parte de sus enlaces y al estar conformada por una gran cantidad de moléculas, los valores registrados en el espectro FTIR contemplan los mismos enlaces de todas las moléculas que constituyen la muestra que se analiza, tal es el caso de las proteínas, en la región 1700-1600  $\text{cm}^{-1}$  (50.96472 a 47.96679 THz) del espectro FTIR

se registran vibraciones principalmente producidas por el doble enlace carbono oxígeno (C=O) atribuidos a las proteínas (específicamente a la amida I), sin embargo, si la muestra que se analiza está constituida por más de un tipo de proteínas, los enlaces C=O de todas ellas aportan a la medida registrada en el espectro FTIR en esa región, además, si otro componente de la muestra también tiene ese enlace de igual manera se registrará en el espectro FTIR [13, 16].

Lo anterior ha dificultado la adopción de una metodología fiable para analizar espectros de muestras biológicas complejas con fines de apoyar en el diagnóstico clínico. En [15] se recopilan un conjunto de técnicas de clasificación/regresión que han permitido obtener buenos resultados en la caracterización de poblaciones a partir de espectros FTIR, las técnicas mencionadas son K-Vecinos más cercanos (K-NN), Análisis Discriminante Lineal (LDA), Modelos de regresión multi-variable (MLRM), Máquinas de Soporte Vectorial (SVM) y Redes Neuronales Artificiales (ANN). Debido a que el objetivo del presente trabajo fue analizar la posibilidad de estimar una variable numérica (valor de hemoglobina glicosilada) no categórica, se evaluó el desempeño de modelos de los modelos de regresión MLRM, SVM y ANN.

La estructura de la ANN con la que se experimentó corresponde a la estructura más simple como se sugiere en [17] con la finalidad de tener una perspectiva general de esta metodología y posteriormente modificarla incluyendo más capas o neuronas. La configuración consistió de una sola capa con una neurona que implementa una función de activación del tipo tangente hiperbólica.

La SVM con la que se experimentó implementa una kernel polinomial de grado 2 con un costo penalización (C) por violación de restricciones de 1 y un valor  $\xi$  de la función de pérdida de 0.1. El MLRM es similar al modelo de regresión simple con la salvedad de que considera dos o más variables independientes (todos los valores del espectro FTIR para el presente caso de estudio). Para evaluar los modelos se emplearon las técnicas de validación cruzada *Leave-One-Out* (LOOCV) y Hold out con una distribución 80-20 (para esta última), es decir, el 80% de las muestras se emplearon en el entrenamiento y 20% se usaron para el proceso de evaluación.

### 3. Resultados

El género, edad promedio y promedio de años de la evolución de la diabetes de los pacientes participantes se presentan en la Tabla 1. Los valores de hemoglobina registrados por los pacientes oscilan entre 5 y 11 mg/dl de acuerdo a los registros de la UEM. En la Fig. 1, se proporciona un histograma con la distribución de los valores registrados, en ella se puede apreciar que la mayoría de los pacientes presentaron valores de entre 5.5 y 6.0 mg/dl valor que de acuerdo con lo señalado por la ADA corresponde a personas diabéticas controladas.

Para la realización del estudio se agruparon los valores de hemoglobina en incrementos de 0.5 mg/dl, así, por ejemplo, el grupo 1 considera espectros de personas que registraron valores de entre 5 y 5.5 mg/dl. Los grupos quedaron conformados como se indica en la Tabla 2. En la Fig. 2 se presenta un espectro promedio de cada uno de los 12 grupos conformados. En la Fig. 3, se presentan los resultados obtenidos mediante un diagrama de cajas con la finalidad de identificar si las metodologías generan algún valor atípico y evaluar el comportamiento de las salidas de los modelos de regresión

con la finalidad seleccionar el que permita obtener la mayor distancia entre grupos con rangos lo mejor posible definidos para los valores de salida. Con fundamento en la Fig 3. es posible determinar que considerando el modelo de segmentación LOOCV de la base de datos para los procesos de entrenamiento-evaluación, los mejores resultados se obtienen mediante ANN, ya que las salidas del modelo presentan rangos definidos y únicos para los 12 grupos de valores de hemoglobina, adicionalmente podemos inferir que los resultados obtenidos presentan una fuerte correlación inversa ya que el grupo 1 que contempla espectros con los menores valores de hemoglobina registrados por los pacientes (5-5.5 mg/dl) producen los valores de salida más altos mientras que el grupo 12 que contempla espectros con los mayores valores de hemoglobina registrados por los pacientes (10.6-10.9 mg/dl) registran los valores de salida más bajos.

Puntualmente los rangos de los valores de salida de ANN se presentan en la Tabla El porcentaje correlación obtenido tras comparar los valores de salida de ANN con los valores reales de hemoglobina registrados por las pacientes fue de -99%, lo cual confirma la fuerte correlación inversa; el grado de relación permite inferir que si se incorpora un modelo de regresión lineal simple (LM) sería posible tener un valor aproximado al valor definido de los grupos (de 1 a 12). Los resultados de evaluar mediante un modelo de regresión las salidas de ANN y los valores de los conjuntos definidos para agrupar los distintos valores de hemoglobina (de 1 a 12) se presentan en la Fig. 4, en ella es posible apreciar que los valores estimados mediante ANN+LM se aproximan de buena manera a los valores reales asignados.

A través del modelo ANN+LM fue posible obtener un coeficiente de determinación ( $r^2$ ) de 0.99 y un valor para la raíz del error cuadrático medio (RMSE) de 0.055 lo que sugiere que considerando la metodología LOOCV, es posible asociar espectros FTIR de muestras de saliva con valores de hemoglobina de pacientes diabéticos. Además de no requerir ningún reactivo para el preprocesamiento de la muestra, el tiempo de procesamiento de las 258 muestras mediante ANN+LM fue de aproximadamente 125 segundos lo que indica que la presente metodología podría ser empleada de buena manera para realizar este estudio de una manera más ágil y a menor costo que la metodología tradicional basada en sangre.

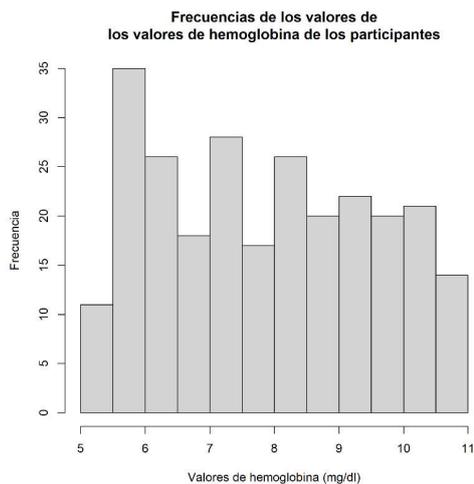
A pesar de que es posible encontrar diversos trabajos que emplean el modelo LOOCV para evaluar las técnicas de clasificación /regresión, en trabajos como los de [15] sugieren que los resultados obtenidos mediante LOOCV se deben al denominado *overfitting* de los métodos, señalando incluso que LOOCV resulta conveniente con una base de datos de hasta 20 muestras.

Por otra parte, trabajos como los de [18, 19] señalan que LOOCV resulta conveniente cuando se cuenta con una cantidad de muestras inferior a la cantidad de variables que componen a cada señal, tal es el caso del presente trabajo ya que se cuenta con un total de 258 espectro y cada espectro se componen de 3736 variables.

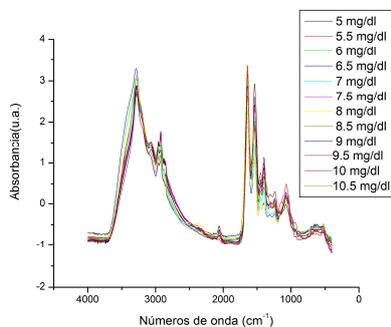
Con la finalidad de evaluar la metodología propuesta en el presente trabajo, se segmentó la base de datos considerando una distribución 80-20 (hold out) para realizar los procesos de entrenamiento y evaluación respectivamente, de esta forma el subconjunto de entrenamiento quedó conformado por 206 espectros y el subconjunto de evaluación por 52, los resultados se presentan en la Fig. 5, en ella se puede apreciar que a diferencia de la metodología LOOCV, con hold out el traslape entre clases aparece para todos los modelos de regresión. Los valores de  $r^2$  y RMSE obtenidos mediante hold out se presentan en la Tabla 4.

**Tabla 1.** Información de los pacientes analizados.

Género	Edad promedio (años)	Tiempo de evolución de la enfermedad (años)
Masculino	119	60±11
Femenino	139	11±8



**Fig. 1.** Distribución de los valores de hemoglobina de los pacientes participantes.



**Fig. 2.** Espectros promedio de cada una de las poblaciones conformadas a partir de las 258 muestras.

Con fundamento en la Fig. 5 y Tabla 4 puede señalarse SVM como la mejor opción considerando hold out. La necesidad del constante monitoreo del nivel de glucosa en sangre de personas diagnosticadas con diabetes ha permitido que distintos trabajos de investigación encaminen sus objetivos a proponer metodologías más accesibles y/o que ocasionen una menor incomodidad física del paciente.

En el presente trabajo se evaluó la viabilidad de emplear espectros FTIR de muestras de saliva con la finalidad de estimar los valores de hemoglobina de un paciente diabético. El uso de espectroscopia FTIR con la finalidad de estimar los valores de

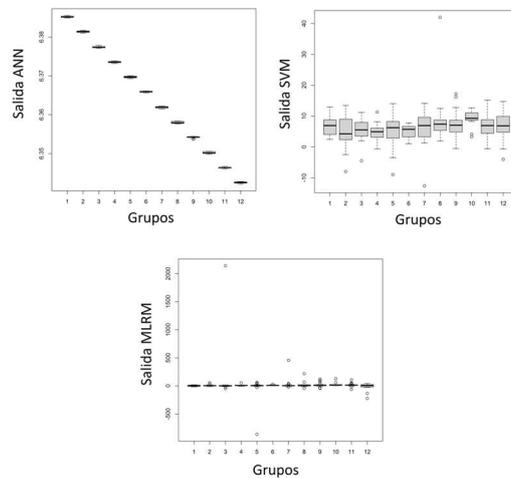


Fig. 1. Salidas obtenidas de los modelos evaluados considerando LOOCV.

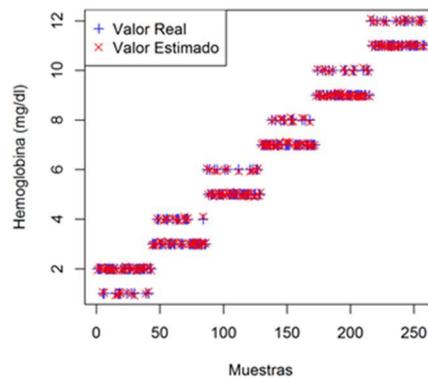


Fig. 4. Comparación de valores de reales de hemoglobina contra los valores estimados mediante un modelo ANN+LM.

hemoglobina de pacientes diabéticos ha sido abordado por trabajos como los de [20]–[23] con particulares diferencias con respecto al objetivo del presente trabajo.

En [20] se propone una metodología basada en regresión lineal que permite obtener una  $r^2$  de 0.99 lo que sugiere que es una técnica muy exacta, sin embargo, los autores analizan una solución artificial no un fluido corporal; al tener control sobre los componentes con los que se realiza esta solución, las variaciones en los espectros FTIR estarán en función de la cantidad de glucosa añadida a cada muestra de la solución y el comportamiento de los fluidos del cuerpo humano presentan una variación más compleja dependiendo de las condiciones actuales del paciente, por lo que el modelo de regresión propuesto en [20] pudiera presentar complicaciones al momento de

**Tabla 2.** Agrupación de espectros en función de niveles de hemoglobina de los pacientes.

Grupo	Rango de glucosa (mg/dl)	Espectros del grupo	Grupo	Rango de glucosa (mg/dl)	Espectros del grupo
1	5-5.5	11	7	8-8.5	32
2	5.6-5.9	32	8	8.6-8.9	11
3	6-6.5	29	9	9-9.5	31
4	6.6-6.9	14	10	9.6-9.9	12
5	7-7.5	32	11	10-10.5	29
6	7.6-7.9	11	12	10.6-10.9	14

**Tabla 1.** Rangos de salida de ANN.

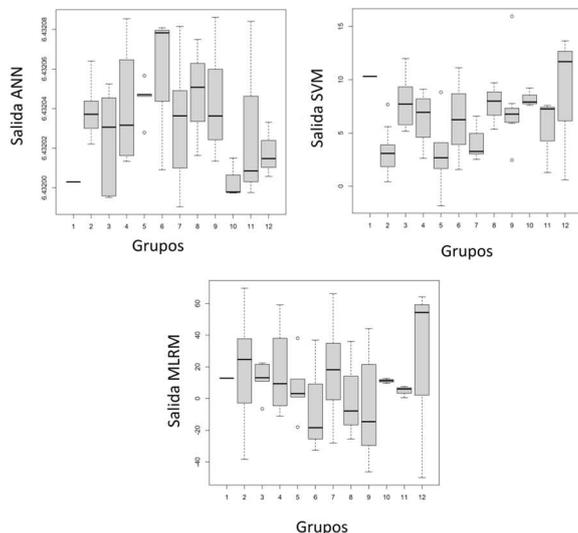
Grupo	Rango de Glucosa (Mg/Dl)	Espectros del Grupo	Rangos de Salida
1	5-5.5	11	6.384928–6.385555
2	5.6-5.9	32	6.381706-6.380961
3	6 - 6.5	29	6.377074–6.377816
4	6.6-6.9	14	6.373185-6.373867
5	7-7.5	32	6.369281-6.370033
6	7.6-7.9	11	6.365584-6.366089
7	8-8.5	32	6.361489-6.362205
8	8.6-8.9	11	6.357612-6.358347
9	9-9.5	31	6.353717-6.35441
10	9.6-9.9	12	6.349809-6.350548
11	10-10.5	29	6.346014-6.346679
12	10.6-10.9	14	6.342136-6.342791

evaluarlo con muestras reales. Los autores de [21] también emplean un modelo de regresión para estimar los valores de hemoglobina de pacientes, pero a través de espectros obtenido de la piel y labios de pacientes diabéticos. El modelo de [21] permitió obtener una  $r^2$  de 0.62 en una población de 28 pacientes.

Uno de los principales inconvenientes señalado por los autores de [21] es la metodología que emplearon para estimar su  $r^2$  ya que es una técnica propuesta por ellos, no la metodología tradicional, aunque reconocen que, para emplear su método en un ambiente real, su técnica de estimación de hemoglobina debe evaluarse apropiadamente siguiendo una metodología convencional, lo anterior impide contrastar los resultados obtenidos en el presente trabajo con [21].

En [22] se propone un modelo de regresión por mínimos cuadrados parciales para estimar los valores de hemoglobina de 232 muestras de sangre de distintos pacientes, si bien el modelo empleado permite obtener una  $r^2$  de 0.96, los autores requirieron de hemolizar la muestra para facilitar el análisis de las mismas pese a que en la sangre es posible encontrar hemoglobina a diferencia del bio fluido (saliva) empleado en el presente proyecto.

Aunque la técnica propuesta en [22] podría agilizar la estimación de los valores de hemoglobina, la necesidad de usar reactivos para el pre procesamiento de la muestra y el malestar físico ocasionado a los pacientes por la extracción de sangre permanece constante. En [23], de manera similar a [22], se analizaron 232 muestras de sangre,



**Fig. 2.** Resultados obtenidos mediante los modelos regresión considerando la metodología hold out 80-20.

**Tabla 2** Métricas obtenidas a partir de los modelos ANN, SVM y MLRM considerando la metodología hold out.

Método	r <sup>2</sup>	RMSE
ANN	0.007	3.28
SVM	0.055	4.25
MLRM	0.013	29.47

pero empleando un modelo de regresión lineal. A través de su modelo reportan una r<sup>2</sup> de 0.94. Para el estudio de [23], las muestras también se hemolizaron lo que ocasiona los mismos inconvenientes que los señalados para [22].

A pesar de que los resultados obtenidos en el presente trabajo empleando hold out son menores a los resultados expuestos por [22, 23] la naturaleza del fluido que se analiza resulta considerablemente distinta, en nuestro caso, se analiza saliva, fluido en el que no hay presencia de hemoglobina, además, la muestra analizada no es sometida a ningún pre procesamiento con algún reactivo y la cantidad de muestras analizada en el presente estudio es superior.

Aunado a lo anterior, siguiendo la metodología de validación cruzada LOOCV fue posible proponer una técnica basada en redes neuronales y regresión lineal simple que permite obtener un valor de r<sup>2</sup> superior a los trabajos relacionados.

#### 4. Conclusiones y trabajo a futuro

En el presente trabajo se evaluó la posibilidad de asociar espectros FTIR con valores de hemoglobina glicosilada en pacientes diabéticos. Se evaluaron los modelos de

regresión configurados a partir de ANN, SVM y MLRM mediante las metodologías de validación cruzada LOOCV y Hold out con una distribución 80-20.

Considerando la metodología LOOCV fue posible conseguir un valor de  $r^2$  de 0.99 y un RMSE de 0.055 mg/dl acoplando un ANN y LM para la estimación de los valores de hemoglobina de un paciente a partir del análisis de espectros FTIR de muestras de saliva, estos valores sugieren que la metodología propuesta resulta mejor opción que las propuestas por trabajos similares, sin embargo, al emplear una segmentación hold out los mejores resultados fueron obtenidos mediante SVM permitiendo obtener valores de 0.055 y 4.25 para  $r^2$  y RMSE respectivamente.

A partir de estos valores concluimos que la metodología propuesta resulta viable, pero es necesario recabar un mayor número de muestras para poder obtener valores similares al modelo ANN+LM, pero con una segmentación hold out de la base de datos.

En este mismo sentido, es necesario experimentar con técnicas de pre procesamiento y selección de atributos con la finalidad de reducir la cantidad de números de onda o frecuencias requeridas para segmentar las poblaciones no sólo para reducir el tiempo de procesamiento sino para poder identificar las frecuencias (números de onda) fundamentales y así construir un dispositivo portable que emita únicamente esas frecuencias y no todas las de la región media del espectro IR, este dispositivo evaluaría los valores obtenidos median ANN+LM (o el método más adecuado) y permitiría estimar los valores de hemoglobina del paciente usando saliva, esta medición resulta más relevante que los valores que se obtienen mediante los glucómetros tradiciones en donde las mediciones de glucosa reflejan únicamente el estado actual del paciente y no un historial de entre 2 y 3 meses como los valores de hemoglobina.

## Referencias

1. Diabetes [Internet]. [cited 2021 Mar 22]. Available from: <https://www.who.int/news-room/fact-sheets/detail/diabetes>
2. Diagnosis | ADA [Internet]. [cited 2021 Mar 22]. Available from: <https://www.diabetes.org/a1c/diagnosis>
3. Whitley, H. P., Yong, E. V., Rasinen, C.: Selecting an A1C Point-of-Care Instrument. *Diabetes Spectrum*, vol 28, no. 3, pp. 201–208 (2015) doi: 10.2337/diaspect.28.3.201
4. Paluszkiwicz, C., Pięta, E., Woźniak, M., Piergies, N., Koniewska, A., Ścierański, W., Misiołek, M., Kwiatek, W. M.: Saliva as a first-line diagnostic tool: A spectral challenge for identification of cancer biomarkers. *Journal of Molecular Liquids*, vol. 307 (2020) doi: 10.1016/j.molliq.2020.112961
5. Saitou, M., Gaylord, E. A., Xu, E., May, A. J., Neznanova, L., Nathanm S., Grawe, A., Chang, J., Ryan, W., Ruhl, S., Knox, S. M., Gokcumen, O.: Functional specialization of human salivary glands and origins of proteins intrinsic to human saliva. *Cell Reports*, vol. 33, no. 7, pp. 108402 (2020) doi: 10.1016/j.celrep.2020.108402
6. Miller, C. S., Foley, J. D., Bailey, A. L., Campell, C. L., Humphries, R. L., Christodoulides, N., Floriano, P. N., Simmons, G., Bhagwandin, B., Jacobson, J. W., Redding, S. W., Ebersole, J. L., McDevitt, J. T.: Current developments in salivary diagnostics. *Biomarkers in Medicine*, vol. 4, no. 1, pp. 171–89 (2010)doi: 10.2217/bmm.09.68
7. Mirzaii-Dizgah, M. H., Mirzaii-Dizgah M. R., Mirzaii-Dizgah, I.: Serum and saliva total tau protein as a marker for relapsing-remitting multiple sclerosis. *Medical Hypotheses*, vol. 135, pp. 109476 (2020) doi: 10.1016/j.mehy.2019.109476
8. Smith, B. C.: *Fundamentals of Fourier transform infrared spectroscopy*. CRC Press (2011)
9. Smith, B. C.: *Infrared spectral interpretation: A systematic approach*. CRC Press (2018)

10. Mid-IR Spectroscopy. ScienceDirect, [Internet]. [cited 2021 Mar 22]. Available from: <https://www.sciencedirect.com/topics/chemistry/mid-ir-spectroscopy>
11. Shi, J., Wong, T. T. W., He, Y., Li, L., Zhang, R., Yung, C. S., Hwang, J., Maslov, K., Wang, L. V.: High-resolution, high-contrast mid-infrared imaging of fresh biological samples with ultraviolet-localized photoacoustic microscopy. *Nature Photonics*, vol. 13, no. 9, pp. 609–615 (2019) doi: 10.1038/s41566-019-0441-3
12. Türker-Kaya, S., Huck, C. W.: A review of mid-infrared and near-infrared imaging: Principles, concepts and applications in plant tissue analysis. *Molecules*, vol. 22, no. 1, pp. 168 (2017) doi: 10.3390/molecules22010168
13. Bel'skaya, L. V., Sarf, E. A., Makarova, N. A.: Use of fourier transform IR spectroscopy for the study of saliva composition. *Journal of Applied Spectroscopy*, vol. 85, no. 3, pp. 445–451 (2018) doi: 10.1007/s10812-018-0670-0
14. Takamura, A., Watanabe, K., Akutsu, T., Ozawa, T.: Soft and robust identification of body fluid using fourier transform infrared spectroscopy and chemometric strategies for forensic analysis. *Scientific Reports*, vol. 8 (2018) doi: 10.1038/s41598-018-26873-9
15. Morais, C. L. M., Lima, K. M. G., Singh, M., Martin, F. L.: Tutorial: Multivariate classification for vibrational spectroscopy in biological samples. *Nature Protocols*, vol. 15, no. 7, pp. 2143–2162 (2020) doi: 10.1038/s41596-020-0322-8
16. Naseer, K., Ali, S., Qazi, J.: ATR-FTIR spectroscopy as the future of diagnostics: A systematic review of the approach using bio-fluids. *Applied Spectroscopy*, vol. 56, no. 2, pp. 85–97 (2020) doi: 10.1080/05704928.2020.1738453
17. Lantz, B.: *Machine Learning with R: Expert techniques for predictive modeling*. Packt Publishing (2013)
18. Liò, P., Zuliani, P.: *Automated reasoning for systems biology and medicine*. Springer, vol. 30 (2019)
19. Andrade-Garda, J.: *Basic chemometric techniques in atomic spectroscopy*. Royal society of chemistry (2013)
20. Shan, X., Chen, L., Yuan, Y., Liu, C., Zhang, X., Sheng, Y., Xu, F.: Quantitative analysis of hemoglobin content in polymeric nanoparticles as blood substitutes using Fourier transform infrared spectroscopy. *Journal of Materials Science*, vol. 21, no. 1, pp. 241–249 (2010) doi: 10.1007/s10856-009-3864-4
21. Yoshida, S., Yoshida, M., Yamamoto, M., Takeda, J.: Optical screening of diabetes mellitus using non-invasive Fourier-transform infrared spectroscopy technique for human lip. *Journal of Pharmaceutical and Biomedical Analysis*, vol. 76, pp. 169–176 (2013) doi: 10.1016/j.jpba.2012.12.009
22. Pan, T., Li, M., Chen, J., Xue, H.: Quantification of glycosylated hemoglobin indicator HbA1c through near-infrared spectroscopy. *Journal of Innovative Optical Health Sciences*, vol. 7, no. 4, pp. 1350060 (2014) doi: 10.1142/S1793545813500600
23. Han, Y., Pan, T., Zhou, H., Yuan, R.: ATR-FTIR spectroscopy with equidistant combination PLS method applied for rapid determination of glycosylated hemoglobin. *Analytical Methods*, vol. 10, no. 28, pp. 3455–3461 (2018)