

Sistema Bayesiano para la Predicción de la Diabetes

Omar D. Castrillón⁽¹⁾, William Sarache⁽¹⁾ y Eduardo Castaño⁽²⁾

(1) Universidad Nacional de Colombia, Sede Manizales, Facultad de Ingeniería y Arquitectura, Departamento de Ingeniería Industrial, Campus la Nubia, Código Postal 170001, Manizales Colombia. (e-mail: odcastrillong@unal.edu.co; wasarachec@unal.edu.co)

(2) Universidad de Caldas, Facultad de Ciencias para la Salud, Programa de Medicina, Código Postal 17001, Manizales, Colombia. (e-mail: eduardo.castano@ucaldas.edu.co)

Recibido Abr. 20, 2017; Aceptado Jun. 23, 2017; Versión final Jul. 20, 2017, Publicado Dic. 2017

Resumen

En este trabajo se propone un sistema de clasificación Bayesiano para la identificación temprana de la diabetes Mellitus con base en el análisis de algunas variables tales como número de embarazos, presión arterial diastólica, espesor cutáneo del tríceps, índice de masa corporal, herencia y edad. La metodología propuesta define y entrena el sistema propuesto con base en muestras tomadas de pacientes diabéticos y no diabéticos. El sistema se validó con pacientes diferentes, manteniendo la misma proporción entre individuos diabéticos y no diabéticos. Finalmente, el número de aciertos y errores en la detección de esta enfermedad fue comparado contra un test especializado. Los resultados indican que, en el 87,69% de los casos, el clasificador bayesiano logra detectar correctamente esta enfermedad con base en las variables antes mencionadas. No obstante, cuando se agregó la variable “insulina en suero”, el porcentaje aumentó al 98.46%.

Palabras clave: clasificador bayesiano; diabetes; entrenamiento de un sistema; detección automatizada

Bayesian System for Diabetes Prediction

Abstract

This study presents a Bayesian system for the early identification of diabetes Mellitus based on the analysis of some variables such as number of pregnancies, diastolic blood pressure, triceps skin thickness, body mass index, heredity and age. The proposed methodology establishes and trains a Bayesian classification system, based on samples of diabetic and non-diabetic patients. The system was tested with different patients maintaining the same ratio between diabetics and non-diabetic individuals. Finally, to detect this disease, the number of hits and errors obtained by the Bayesian classifier was compared with a specialized test. The results indicate that, considering the aforementioned variables the disease was correctly detected by the system in 87.69% of the analyzed cases. However, when the variable serum insulin was included, this percentage increased up to 98.46%.

Keywords: bayesian classifier; diabetes; training system; automated detection

INTRODUCCIÓN

Según la IDF (International Diabetes Federation, 2014), un gran porcentaje de la población mundial desconoce que padece la diabetes tipo 2. En parte, lo anterior se debe a que su detección se realiza mediante algunos test que no siempre están disponibles en todos los centros de atención de forma oportuna, especialmente en países de bajos ingresos (Raab et al., 2004). Las cifras de la IDF (2013), ubican a Colombia como el segundo país con más ciudadanos diabéticos en la región del Centro y Sur de América. No obstante, a nivel mundial, el número de personas con diabetes se estima en 387 millones de personas cuyas edades oscilan entre los 20 y 79 años. De estos, el 46% aún no está diagnosticado (International Diabetes Federation, 2014).

La diabetes es una enfermedad que genera un gran número de complicaciones tales como retinopatía, problemas renales, hipertensión, problemas cardiovasculares, daños al sistema nervioso, entre otras que, incluso, pueden conducir a la muerte (National Center for Chronic Disease Prevention and Health Promotion, 2010). Actualmente, su detección se basa en tres métodos que miden el nivel de azúcar en la sangre: a) prueba de glucosa sanguínea en ayunas; b) examen de tolerancia oral a la glucosa; y c) valoración de la hemoglobina glicosilada, la cual permite conocer el comportamiento del azúcar en la sangre en los últimos meses. El problema de tales métodos radica en que requieren pruebas de laboratorio, las cuales pueden tardar un par de días, o requerir el empleo de algunos equipos instantáneos de medición que pueden ser costosos o no encontrarse a disposición de toda la población, especialmente en comunidades rurales o en países de bajo desarrollo en los cuales la cobertura en el servicio de salud es deficiente.

En atención a esta problemática, el presente estudio propone un sistema predictivo para diagnosticar la diabetes basado en un clasificador Bayesiano. Las redes bayesianas son un modelo probabilístico por medio del cual es factible construir un grafo entre las causas de un evento (variables independientes) y sus consecuencias (Variables dependientes) (Mesa et al., 2011). Los clasificadores bayesianos permiten clasificar eventos discretos y limitados (variables independientes) en un número determinado de clases (Hernández et al 2010), definiendo una función estadística para cada clase. En la definición de estas funciones estadísticas toman como referencia una base de datos de entrenamiento. Con base en estas funciones definidas, el sistema podrá clasificar un nuevo conjunto de variables independientes (datos de prueba) y establecer la clase a la que pertenecen, con base en la función estadística que genera el mayor valor. En el caso de enfermedades, el sistema opera a partir de una serie de parámetros simples tomados del paciente y no exige análisis de laboratorio.

Al revisar las contribuciones que han abordado el desarrollo de sistemas inteligentes aplicados a la detección de la diabetes en los últimos 20 años, se encontró que, si bien se han diseñado algunos sistemas de detección basados en técnicas convencionales (Torres-Zapata et al., 2012) y otros desde la perspectiva de los sistemas inteligentes (Herrero et al, 2017; Kavakiotis et al, 2017; Gargeya et al, 2017), estos no han sido exploradas en su totalidad. En la Tabla 1 se relacionan algunos estudios identificados.

Tabla 1: Algunas contribuciones sobre sistemas inteligentes aplicados a la detección de la diabetes

<i>Autores</i>	<i>Contribución</i>
Ping (1997)	Propone una nariz electrónica para la detección de la diabetes, basada en la mimetización del sistema olfatorio. Reporta buenos resultados.
Polat y Güneş (2007)	Propone un sistema experto basado en un análisis componentes principales (PCA) y un sistema de neuro- difuso. Nivel de acierto: 89.47%
Kahramanli y Allahverdi (2008)	Emplearon un sistema de clasificación por medio de una red neuronal y una red neuronal difusa. Nivel de acierto: 84.24% a 86.8%.
Lekkas y Mikhailov (2010)	Diseñaron una metodología de clasificación difusa con dos propósitos; el primero, para la clasificación de diabetes con resultados de aciertos en la detección de la enfermedad cercanos al 92% y, el segundo, en dermatología. Nivel de acierto: 97.55%.
Dogantekin et al. (2010)	Propusieron un análisis discriminante lineal para la detección de la diabetes y un sistema de inferencia difuso basado en una red adaptativa. Se empleó un conjunto de variables similares a las empleadas en este trabajo y otras adicionales. Nivel de acierto: 84.61%.
Ganji y Abadeh (2011)	Emplea un sistema de clasificación basado en colonia de hormigas para extraer un conjunto de reglas el cual permita diagnosticar la diabetes. Se emplean las mismas características trabajadas en este artículo y otras adicionales. Nivel de acierto: 84.24%.
Karan et al. (2012)	Diagnostica la diabetes mediante una red neuronal y un sistema computacional de cuidado de la salud, en un esquema cliente servidor. El sistema es entrenado con algunos parámetros similares a los manejados por este documento. Con una tasa de aprendizaje del 95%, el sistema presenta un error cuadrático del 4.86%. El modelo puede ser extendido a otras enfermedades.

Tabla 1 (continuación)

Khan (2013)	Diseñó un sistema de clasificación de diabetes basado en el algoritmo de Levenberg-Marquardt (LM). Este algoritmo es aplicado sobre una red neuronal, entrenando el sistema permanentemente hasta alcanzar un óptimo. Nivel de acierto: 90% cuando no se tiene la enfermedad y un porcentaje más bajo cuando se tiene la enfermedad.
Aslam et al. (2013)	Emplearon un algoritmo de programación genética en el diagnóstico de la diabetes. La clasificación se realiza con la ayuda de una máquina de soporte vectorial y Redes Neuronales. Se emplean las mismas variables de clasificación de este documento y otras adicionales. Nivel de acierto: 87%.
Beloufa y Chikh (2013)	Diagnosticaron la diabetes por medio de una colonia de abejas artificial con operadores de mutación, proponiéndose un clasificador basado en lógica difusa. Se emplearon las mismas variables de clasificación del presente artículo y otras adicionales. Nivel de acierto: 84.21%.
Lai et al. (2013)	Se detecta la enfermedad por medio de un sistema de inferencia con evolución diferencial de la transformada de Wavelet. Nivel de acierto: 90.07%.
Lai et al. (2014)	Se propone un sistema de interferencia difuso, optimizado bajo un sistema inteligente, para analizar la relación de la hipoglicemia y diferentes parámetros psicológicos, en pacientes con diabetes tipo I. Nivel de acierto: 90.10%.
Lukmanto y Irwansyah (2015)	Emplearon un modelo jerárquico difuso en la detección de la diabetes. Nivel de acierto: 87.46%
Santhanam y Padmavathi (2015)	Propusieron un sistema de detección de diabetes el cual inicia por medio de una remoción del ruido de los datos, los cuales posteriormente son clasificados por medio de una máquina de soporte vectorial. Nivel de acierto 96.71%.
Zhu et al. (2015)	Usando una combinación ponderada de múltiples factores, diseñaron un sistema basado en múltiples clasificadores para mejorar la detección de la diabetes. Este sistema emplea algunos de los parámetros trabajados en este artículo y otras adicionales. Nivel de acierto: 90%.
Erkaymaz y Ozer (2016)	Propusieron un sistema de clasificación de diabetes basado en una red neuronal artificial de 9 neuronas en capas ocultas. Este trabajo empleó las mismas variables de clasificación trabajadas en este artículo y otras adicionales. Nivel de acierto: 91.6%.
Amina et al (2017)	Se propone un sistema de clasificación automática para identificar lesiones de renopatía producto de la diabetes. Nivel de acierto: 91.6%. 98%.
Kavakiotis et al. (2017)	Se realiza un análisis sobre las técnicas más empleadas en los procesos de predicción, diagnóstico, complicaciones, antecedentes y gestión de la diabetes.

Como se puede observar en la Tabla 1, predomina el uso de técnicas difusas y el uso de algoritmos bioinspirados. Aunque los clasificadores Bayesianos han sido usados en problemas relacionados (Magni et al., 2010), así como en algunos modelos matemáticos (Klinke, 2007; Morbiducci et al., 2014; Brubaker et al., 2007), no se encontraron trabajos centrados en la identificación temprana de enfermedades como la diabetes. Por otro lado, aunque el clasificador Bayesiano propuesto logra resultados similares a los obtenidos por otros autores, cuando la variable *insulina en suero* fue incorporada, la capacidad de predicción mostró resultados muy superiores a los reportados en la literatura.

Para su presentación, el artículo ha sido estructurado de la siguiente manera: en primer lugar se explica la metodología de cuatro pasos utilizada para construir el clasificador Bayesiano. Seguidamente, se exponen y discuten los resultados obtenidos utilizando información de pacientes con y sin diabetes. Finalmente, se presentan las conclusiones más relevantes y algunas líneas de investigación que podrían ser exploradas.

MATERIALES Y MÉTODOS

La metodología utilizada está compuesta por los siguientes pasos: 1) Definición de las bases de datos; 2) Estructuración matemática del clasificador bayesiano; 3) Construcción del algoritmo de clasificación; y 4) Validación del sistema.

Paso 1. Definición de las bases de datos. Se definieron cuatro bases de datos, dos de entrenamiento y otra dos de prueba, según se expone en la Tabla 2. En esta tabla, cada columna representa una característica del paciente objeto de análisis. Las ocho características analizadas fueron: número de embarazos (C_1); glucosa en plasma (C_2); presión sanguínea diastólica (C_3); espesor del pliegue cutáneo del tripces (C_4); insulina en suero (C_5); índice de masa corporal (C_6); función de pedigrí de la diabetes (C_7); edad (C_8). Una de las bases de datos de entrenamiento corresponde a personas que padecen la enfermedad y la otra a personas que no la padecen. Estas se construyeron tomando como referencia los datos encontrados en el *National Institute of Diabetes and Digestive and Kidney Diseases* (1990). De forma similar, se construyeron las bases de datos de validación.

Tabla 2: Bases de datos de entrenamiento y prueba

Personas	Características								¿Tiene diabetes?
	C ₁	C ₂	C ₃	C ₄	C ₅	C ₆	C ₇	C ₈	
1									
2									
n									

Paso 2. Estructuración matemática del clasificador bayesiano. Se aplicó el siguiente sub - procedimiento: Se define X_i , como una hiper matriz de tres dimensiones que contiene el conjunto de características. Las dimensiones son definidas de la siguiente forma: las filas representan el número de problemas objeto de análisis, las columnas son el número de características definidas en la Tabla 2 (en este caso 8); y c es el número de clases, en este caso 2 (¿tiene diabetes?: Si o No). Seguidamente, se calcula el vector de medias u_i y la matriz de covarianza Σ_i de X_i , dónde i representa la dimensionalidad de cada una de las matrices expresadas en las ecuaciones 1- 4 y K es una constante definida a partir de la matriz X_i . Para cada clase se define una función de probabilidad según el sistema de ecuaciones propuesto por Duda et al., (2001: 41):

$$P_i(x) = X^t W_i X + w_i^t X + w_{i0} \quad (1)$$

Donde:

$$W_i = -\frac{1}{2} \Sigma_i^{-1} \quad (2)$$

$$w_i = \Sigma_i^{-1} u_i \quad (3)$$

$$w_{i0} = -\frac{1}{2} u_i^t \Sigma_i^{-1} u_i - \frac{1}{2} \ln(|\Sigma_i|) + \ln(k) \quad (4)$$

Paso 3. Construcción del algoritmo de clasificación. Tomando como referencia las variables mostradas en la Tabla 2, el algoritmo iniciará un proceso de selección de variables analizando todas las combinaciones posibles como se ilustra en la Tabla 3.

Tabla 3: Posibles combinaciones de variables

Combinaciones	Resultados
Combinatoria (8,1)	8
Combinatoria (8,2)	28
Combinatoria (8,3)	56
Combinatoria (8,4)	70
Combinatoria (8,5)	56
Combinatoria (8,6)	28
Combinatoria (8,7)	8
Combinatoria (8,8)	1
Total	255

Donde la combinatoria (n,k) define todos los posibles subgrupos diferentes de tamaño k , tomados de un grupo mayor de n elementos. Matemáticamente se define como se expresa en la Ecuación 5. Finalmente, tomando como referencia los datos de entrenamiento, y para cada uno de los subgrupos definidos en la Tabla 3, se aplican las ecuaciones 1 a la 4 con el fin de establecer cada una de las clases referidas en la ecuación 1.

$$C(n, k) = \frac{n!}{k!(n-k)!} \quad (5)$$

Paso 4. Validación del sistema. Cada uno de los conjuntos de validación del paso 1, son evaluados en las ecuaciones generadas en los pasos 2 y 3. La ecuación que genere el mayor resultado, supondrá la clase (diabéticos o no diabéticos) a la cual debería pertenecer el paciente analizado. Los resultados generados por el clasificador deben ser comparados con el respectivo diagnóstico médico (diabético o no diabético) para cada paciente, obteniéndose de esta manera el número de aciertos o errores del clasificador.

En síntesis, el sistema de clasificación bayesiano propuesto, trabaja con un conjunto de registros de entrenamiento que permiten definir una función de probabilidad por cada una de las clases de interés

(diabético y no diabético). Posteriormente los registros de pacientes nuevos (diabéticos y no diabéticos) son evaluados en cada una de las funciones de probabilidad definidas de forma *a priori*, suponiéndose que el nuevo paciente pertenece aquella clase cuyo resultado de la función de probabilidad sea máximo. Finalmente, los resultados del clasificador son comparados con los resultados de los test especializados en diabetes, con el fin de establecer el porcentaje de aciertos del sistema.

RESULTADOS Y DISCUSIONES

Paso 1. Definición de las bases de datos. Tomando como referencia las bases de datos del *National Institute of Diabetes and Digestive and Kidney Diseases* (1990) se definieron 4 archivos siguiendo la estructura de la Tabla 1. Cada archivo representa un grupo que consta de 65 registros. Dos grupos de entrenamiento (personas con y sin diabetes) y otros dos grupos de prueba (personas con y sin diabetes). Los parámetros estadísticos básicos de los registros seleccionados por características son ilustrados en las Tablas 4 a la 7.

Tabla 4: Parámetros básicos. Personas sin diabetes-Entrenamiento.

Parámetro	C ₁	C ₂	C ₃	C ₄	C ₅	C ₆	C ₇	C ₈
Promedio	2.88	109.06	67.92	27.22	118.02	31.20	0.44	28.34
Desviación	2.54	26.77	11.13	10.30	88.09	6.82	0.25	8.85
Mínimo	0.00	0.00	30.00	11.00	18.00	19.40	0.09	21.00
Máximo	13.00	180.00	90.00	60.00	485.00	46.80	1.40	61.00
Mediana	2.00	103.00	68.00	27.00	94.00	30.50	0.36	26.00
Varianza	6.45	716.68	123.82	106.08	7759.08	46.57	0.06	78.38

Tabla 5: Parámetros básicos. Personas sin diabetes- Prueba

Parámetro	C ₁	C ₂	C ₃	C ₄	C ₅	C ₆	C ₇	C ₈
Promedio	2.69	114.82	69.94	26.11	143.52	31.61	0.48	28.12
Desviación	2.62	28.62	10.86	11.17	143.26	6.83	0.35	7.35
Mínimo	0.00	71.00	46.00	8.00	37.00	19.30	0.09	21.00
Máximo	12.00	197.00	100.00	52.00	744.00	52.30	2.33	51.00
Mediana	2.00	108.00	70.00	25.00	87.00	31.60	0.42	25.00
Varianza	6.84	819.28	118.00	124.69	20522.10	46.60	0.12	54.02

Tabla 6: Parámetros básicos. Personas con diabetes-Entrenamiento

Parámetro	C ₁	C ₂	C ₃	C ₄	C ₅	C ₆	C ₇	C ₈
Promedio	4.94	139.17	72.22	32.05	198.37	35.81	0.63	36.00
Desviación	4.32	29.28	14.26	9.76	136.01	7.73	0.39	11.02
Mínimo	0.00	78.00	30.00	7.00	29.00	22.90	0.13	21.00
Máximo	17.00	197.00	110.00	56.00	846.00	67.10	2.29	60.00
Mediana	4.00	137.00	70.00	32.00	168.00	34.50	0.55	32.00
Varianza	18.62	857.11	203.48	95.23	18500.05	59.74	0.15	121.34

Tabla 7: Parámetros básicos. Personas con diabetes. Validación o prueba

Parámetro	C ₁	C ₂	C ₃	C ₄	C ₅	C ₆	C ₇	C ₈
Promedio	4.00	151.22	75.94	33.88	215.32	35.75	0.62	35.88
Desviación	3.44	29.39	11.46	9.51	129.80	5.63	0.42	10.33
Mínimo	0.00	80.00	50.00	14.00	14.00	23.40	0.13	21.00
Máximo	12.00	198.00	106.00	63.00	600.00	59.40	2.42	58.00
Mediana	3.00	150.00	78.00	33.00	180.00	35.20	0.53	34.00
Varianza	11.84	863.89	131.25	90.48	16847.66	31.68	0.18	106.61

Los parámetros en cada una de las tablas 4-7 fueron definidos como: C₁= número de embarazos; C₂ = glucosa en plasma; C₃ = presión sanguínea diastólica; C₄= espesor del pliegue cutáneo del tripces; C₅= insulina en suero; C₆= índice de masa corporal; C₇ = función de pedigrí de la diabetes; C₈ = edad.

Paso 2. Estructuración matemática del clasificador bayesiano. Se obtuvieron las funciones de probabilidad para las dos clases analizadas. Estas ecuaciones se definieron de la siguiente manera:

$$P_{\text{normal}}(x) = X_{(1xk)}^t W_{i_{\text{normal}}(kxk)} X_{(kx1)} + w_{i_{\text{normal}}(1xk)} X_{(kx1)} + w_{i_{\text{normal}}} \quad (\text{Sin diabetes}) \quad (6)$$

$$P_{\text{Diabetis}}(x) = X_{(1xk)}^t W_{i_{\text{diabetis}}(kxk)} X_{(kx1)} + w_{i_{\text{diabetis}}(1xk)} X_{(kx1)} + w_{i_{\text{normal}}} \quad (\text{Con diabetes}) \quad (7)$$

Donde, las matrices W_{i_normal} , W_{i_normal} y la constante W_{io_normal} y las matrices $W_{i_diabetes}$, $W_{i_diabetes}$ y la constante $W_{io_diabetes}$, son definidas tomando como referencia las ecuaciones 2 - 4 y el conjunto de datos de entrenamiento establecidos para los pacientes sin diabetes y con diabetes. Adicionalmente, la dimensionalidad de cada una de las matrices empleadas en las ecuaciones 6 y 7, es definida en el paréntesis anexo al lado de cada matriz X , W y w como subíndice; esto es: $(1 \times k)$, $(k \times k)$ y $(k \times 1)$. La dimensionalidad representada por la variable k es definida en el paso 3.

Paso 3. Construcción del algoritmo de clasificación. Para cada uno de los posibles subgrupos de tamaño k (k varía entre 1 y 8) tomados de un grupo mayor n ($n=8$), se definen las ecuaciones de entrenamiento derivadas de las ecuaciones 6 y 7.

Paso 4. Validación del sistema. Tomando como referencia las características empleadas para definir cada una de las ecuaciones del paso 3 (subgrupos de tamaño k tomados de un grupo mayor n), se seleccionaron estas mismas características en cada una de las filas de la matriz que representa el conjunto de datos de validación, tanto para personas diabéticas como para no diabéticas (cada fila representa los datos de un paciente). Los datos seleccionados, representan la variable X en las ecuaciones 6 y 7 definidas en el paso 3. Al resolver las operaciones matriciales para dichas ecuaciones, se obtiene como resultado un número real. En consecuencia, el paciente pertenece a la clase cuyo resultado sea mayor. Esta clasificación es comparada, para cada paciente, con los resultados del test especializado. La Figura 1, muestra la diferenciación entre las clases, definidas por las ecuaciones 6 y 7 e ilustra los mejores resultados de clasificación obtenidos por el sistema para cada uno de los registros validados. El detalle de las k características seleccionadas en el Paso 3, que dieron origen a la Figura 1, es mostrado en la última fila de la Tabla 8. En las columnas 3 y 4, el complemento al 100% de los resultados obtenidos representa los falsos positivos y los falsos negativos respectivamente.

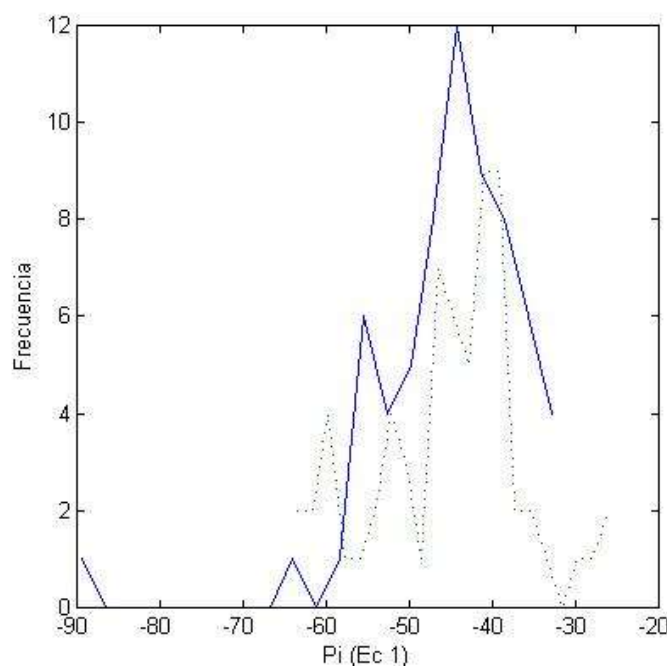


Fig. 1: Diferenciación entre clases. Diabetes línea Sólida. Sin Diabetes Línea Punteada.

Tabla 8: Porcentaje de aciertos del sistema.

Características	Promedio de aciertos	Aciertos pacientes con diabetes	Aciertos pacientes sin diabetes
Nro. embarazos- Presión sanguínea diastólica- Espesor del pliegue cutáneo del tripces - Índice de masa corporal	87.69%	93.84% (4 fallos)	81.54% (12 fallos)
Nro. Embarazos - Glucosa en plasma - Espesor del pliegue cutáneo del tripces - Índice de masa corporal - Edad	93.08%	92.31% (5 fallos)	93.85% (4 fallos)
Presión Sanguínea Diastólica - Espesor del pliegue cutáneo del tripces - Insulina en Suero - Índice de masa corporal - Función de pedigrí de la diabetes - Edad	96.92%	98.46% (1 Fallo)	95.38% (3 fallos)

Los falsos negativos y falsos positivos, (los cuales pueden ser fácilmente calculados como el complemento al 100% de los valores obtenidos en las columnas 3 y 4 respectivamente), muestran valores pequeños, aspecto que establece un alto grado de precisión del sistema.

Los resultados muestran que, a partir de diversas variables empleadas en la literatura tales como el número de embarazos, la presión arterial diastólica, el espesor cutáneo del tríceps, el índice de masa corporal, la herencia y la edad, el porcentaje de aciertos del sistema es bueno (87.69%), en comparación con los trabajos identificados en la literatura. No obstante, cuando la variable *insulina en suero* es incorporada en el estudio, el sistema mejora considerablemente su desempeño (98.46%). Al contrastar este último resultado con las contribuciones que se exponen en la Tabla 1, se encuentra que el sistema propuesto arroja mejores resultados para diagnosticar la diabetes. Este hallazgo facilita aún más la detección de la enfermedad contribuyendo a su prevención.

CONCLUSIONES

Dado los resultados obtenidos, el clasificador bayesiano para identificar personas con diabetes se propone como una herramienta de utilidad para la detección temprana de esta enfermedad, sin recurrir a pruebas de laboratorio. En particular, el nivel de acierto del sistema puede variar entre un 87.6% y un 96.9% en dependencia del número de características del paciente que sean analizadas. Por otro lado, la detección de personas con la enfermedad oscila entre el 92.3% y 98.4%, al tiempo que la detección de personas sin la enfermedad oscila entre el 81.53% y 95.38%. La metodología propuesta está basada en una generación de funciones bayesianas para cada una de las clases creadas. Sobre estas funciones se prueban los datos de entrenamiento y se clasifica el paciente en aquella clase cuya función bayesiana sea máxima. Como futuras líneas de investigación se plantea adaptar el clasificador bayesiano con el fin de que se pueda emplear para detectar tempranamente diversas enfermedades, aspecto que le dará más versatilidad para abordar una mayor diversidad de problemas en el área de la salud.

AGRADECIMIENTOS

Se agradece a la Universidad Nacional de Colombia y en especial al Departamento de Ingeniería Industrial de la sede Manizales.

REFERENCIAS

- Amina, J., M., Sharifa, M., Yasmina, H. A., y S. L. Fernandes, A method for the detection and classification of diabetic retinopathy using structural predictors of bright lesions: *Journal of Computational Science*, 19, 153-164 (2017)
- Aslam, M. W., Z. Zhu y A. K. Nandi, Feature generation using genetic programming with comparative partner selection for diabetes classification, *Expert Systems with Applications*: 40(13), 5402–5412 (2013)
- Beloufa, F. y M. Chikh, Design of fuzzy classifier for diabetes disease using Modified Artificial Bee Colony algorithm, *Computer Methods and Programs in Biomedicine*: 112(1), 92–103 (2013)
- Brubaker, P.L., E.L. Ohayon, L.M. D'alessandro y K.H. Norwich, A Mathematical Model of the Oral Glucose Tolerance Test Illustrating the Effects of the Incretins, *Annals of Biomedical Engineering*: 35(7), 1286–1300 (2007)
- Dogantekin, E., A. Dogantekin, D. Avci y L. Avci, An intelligent diagnosis system for diabetes on Linear Discriminant Analysis and Adaptive Network Based Fuzzy Inference System: LDA-ANFIS, *Digital Signal Processing*: 20(4), 1248-1255 (2010)
- Duda, R., P. E. Hart y D. Stork, *Pattern Classification*. New York: John Wiley, 41 pp. (2001)
- Erkaymaz, O., y M. Ozer, Impact of small-world network topology on the conventional artificial neural network for the diagnosis of diabetes, *Chaos Solitons & Fractals*: 83, 178–185 (2016)
- Ganji, M. F. y M. S. Abadeh, A fuzzy classification system based on Ant Colony Optimization for diabetes disease diagnosis, *Expert Systems with Applications*: 38(12), 14650–14659 (2011)
- Gargeya R. y T. Leng, Automated Identification of Diabetic Retinopathy Using Deep, *Ophthalmology*: 124(7), 962 – 969 (2017)
- Herrero, P., J. Bondia, O. Adewuyi, P. Mohamed, El-Sharkawy, M. Reddy, C. Toumazou, N. Oliver. y P. Georgiou, Enhancing automatic closed-loop glucose control in type 1 diabetes with an adaptive meal bolus calculator – in silico evaluation under intra-day variability, *Computer Methods and Programs in Biomedicine*: 146, 125-131 (2017)

- Hernández, J., M. Castro y M. Angélica, Clasificador Jerárquico de Imágenes utilizando Naive Bayes, Instituto Nacional de Astrofísica Óptica y Electrónica. México (2010)
- International Diabetes Federation, Anual Report 2014, International Diabetes Federation (2014)
- Kahramanli, H. y N. Allahverdi, Design of a hybrid system for the diabetes and heart diseases, *Expert Systems with Applications*: 35(1-2), 82–89 (2008)
- Karan, O., C. Bayraktar, H. Gümüşkaya y B. Karlık, Diagnosing diabetes using neural networks on small mobile devices, *Expert Systems with Applications*: 39(1), 54–60 (2012)
- Kavakiotis, L., O. Tsave., A. Salifoglou, N. Maglaveras, I. Vlahavas y I. Chouvarda, Machine Learning and Data Mining Methods in Diabetes Research, *Computational and Structural Biotechnology Journal*: 15, 104–116 (2017)
- Khan, N., D. Gaurav y T. Kandl, Performance Evaluation of Levenberg-Marquardt Technique in Error Reduction for Diabetes Condition Classification, *Procedia Computer Science*: 18, 2629–2637 (2013)
- Klinke, D.J., Integrating Epidemiological Data into a Mechanistic Model of Type 2 Diabetes: Validating the Prevalence of Virtual Patients, *Annals of Biomedical Engineering*: 36(2), 321–334 (2007)
- Lai, J.C.Y., F.H.F. Leung, S.H. Ling y H.T. Nguyen, Hypoglycaemia detection using fuzzy inference system with multi-objective double wavelet mutation Differential Evolution, *Applied Soft Computing*: 13(5), 2803–2811 (2013)
- Lai, J.C.Y., F.H.F. Leung y S.H. Ling, Hypoglycaemia detection using fuzzy inference system with intelligent optimizer, *Applied Soft Computing*: 20, 54–65 (2014)
- Lekkas, S. y L. Mikhailov, Evolving fuzzy medical diagnosis of Pima Indians diabetes and of dermatological diseases, *Artificial Intelligence in Medicine*: 50(2), 117–26 (2010)
- Lukmanto, R.B. y E. Irwansyah, The Early Detection of Diabetes Mellitus (DM) Using Fuzzy Hierarchical Model, *Procedia Computer Science*: 59, 312–319 (2015)
- Magni P., R. Bellazzi, G. Sparacino y C. Cobelli. Bayesian Identification of a Population Compartmental Model of C-Peptide Kinetics, *Annals of Biomedical Engineering*: 28, 812–823 (2000)
- Mesa, L., M. Lozano y J. Romero, Descripción general de la Inferencia Bayesiana y sus aplicaciones en los procesos de gestión, *La simulación al servicio de la academia*: 2, 1-28 (2011)
- Morbiducci, U., G. Di benedetto, L. Gaetano, A. Kautzky-willer, G. Paccini y A. Tura, Redicting the Metabolic Condition After Gestational Diabetes Mellitus from Oral Glucose Tolerance Test Curves Shape, *Annals of Biomedical Engineering*: 42(5), 1112- 1120 (2014)
- National Center for Chronic Disease Prevention and Health Promotion, Controle su diabetis, Guía para el cuidado de su Salud: Atlanta, Georgia, Estados Unidos (2010)
- National Institute of Diabetes and Digestive and Kidney Diseases. Pima Indians Diabetes Database. 1990. Recuperado de: <https://goo.gl/1Umksp> y <https://goo.gl/DweQ8k>. Febrero (2016)
- Ping, W., T. Yi, X. Haibao y S. Farong, A novel method for diabetes diagnosis based on electronic nose, *Biosensors and Bioelectronics*: 12(9), 1031–1036 (1997)
- Polat, K. y S. Güneş, An expert system approach based on principal component analysis and adaptive neuro-fuzzy inference system to diagnosis of diabetes disease, *Digital Signal Processing*: 17(4), 702–710 (2007)
- Raab R., L., Fezeu y J, Mbanya, Coste y disponibilidad de suministros diabéticos: encuesta de la FID 2002-2003, *Dibaetis Voice*: 49(3), 24-27 (2004)
- Santhanam, T. y M.S. Padmavathi, Application of K-Means and Genetic Algorithms for Dimension Reduction by Integrating SVM for Diabetes Diagnosis, *Procedia Computer Science*: 47, 76–83 (2015)
- Torres-Zapata, Á.E., M.A. Aparicio-Trápala, J.L. Blé Castillo, C.A. Corzo-Sosa, Respuesta Glucémica e Insulínica de Pacientes con Diabetes Tipo 2 al consumo de Sopa de Calabaza Criolla (*Cucúrbita Pepo L.*) Enriquecida con Almidón de Banano, *Información Tecnología*: 23(2), 71-86 (2012)
- Zhu, J., Q. Xie y K. Zheng, An improved early detection method of type-2 diabetes mellitus using multiple classifier system, *Information Sciences*: 292,1–14 (2015)