

## **Mecanismo de clasificación para diabetes mellitus en la población de Aguascalientes, México**

Pablo Rodríguez de León, María Dolores Torres Soto, Aurora Torres Soto

Benemérita Universidad Autónoma de Aguascalientes,  
Posgrado e Investigación,  
México

pablo.rdz@live.com, mdtorres@correo.uaa.mx,  
atorres@correo.uaa.mx

**Resumen.** Año con año en México, se pierden muchas vidas debido a diabetes mellitus la cual es una enfermedad crónica asociada a otras enfermedades crónicas como hipertensión, dislipidemia y síndrome metabólico, por esta razón nació la inquietud de identificar tempranamente las variables características relacionadas al desarrollo de esta enfermedad. Esta investigación presenta la creación de un mecanismo de clasificación para seleccionar subconjuntos de características y clasificar pacientes con la patología de diabetes mellitus en la población de Aguascalientes, México. En esta investigación se diseñó, desarrollo y puso a punto un mecanismo integrado por dos mecanismos que fueron ejecutados de forma secuencial, el primero fue utilizando la técnica de selección de variables de testores típicos, con este mecanismo se obtuvieron los sub conjuntos de características interrelacionadas que mejor describen un paciente con la patología de diabetes mellitus, en el segundo mecanismo se empleó la técnica de redes neuronales artificiales con la cual se diseñó, desarrollo, entreno y se validó un clasificador con una precisión de casi el 92%. Cabe mencionar que en esta investigación se contó con información extraída de consultas de pacientes del sistema expediente clínico electrónico del ISSEA, el cual basa sus diagnósticos en el estándar internacional CIE 10. Como resultados de esta investigación se desarrolló un clasificador con una precisión de casi el 92% para identificación de pacientes con diabetes mellitus en la población de Aguascalientes, y por medio de la selección de variables se identificaron variables que no son comúnmente consideradas en conjunto en la literatura, tales como el estado civil, existencia de complicaciones y antecedentes familiares con diabetes.

**Palabras clave:** Testores típicos, redes neuronales artificiales, diabetes mellitus, catalogo internacional de enfermedades 10.

### **Classification Mechanism for Diabetes Mellitus in the Population of Aguascalientes, Mexico**

**Abstract.** Year after year in Mexico many lives are lost due to diabetes mellitus which is a chronic disease associated with other chronic diseases such as

hypertension, dyslipidemia, and metabolic síndrome, for this reason, the concern was raised to the early identification of the characteristic variables related to the development of this disease. This research presents the creation of a hybrid classification mechanism to select subsets of characteristics and classify patients with the pathology of diabetes mellitus in the population of Aguascalientes, Mexico. In this research, a hybrid mechanism composed of two mechanisms that were executed sequentially was designed, developed, and developed. The first was using the technique of variable selection of Typical Testers, with this mechanism the subsets of interrelated characteristics that best describe a patient with the pathology of diabetes mellitus were obtained. In the second mechanism, the artificial neural network technique was used, a classifier was designed, developed, trained, and validated with an accuracy of almost 92%. It is worth mentioning that in this research, information was obtained from patient consultations of the ISSEA electronic clinical record system, which bases its diagnoses on the international ICD 10 standard. As a result of this research, a classifier was developed with an accuracy of almost 92% for identification of patients with diabetes mellitus in the population of Aguascalientes, and through the selection of variables, variables were identified that are not commonly considered together in the literature, such as the marital status, existence of health complications and family history with diabetes.

**Keywords:** Typical testers, artificial neural networks, diabetes mellitus, international catalog of diseases 10.

## 1. Introducción

En la actualidad, los sistemas de consultas médicas producen una gran cantidad de datos e información, los cuales no son aprovechados y analizados por los organismos que las crean y sustentan, ya que carecen de conocimiento referente a procesamiento, transformación e interpretación de la información.

Para esta investigación se otorgó información de datos de consultas de pacientes de centros de salud del Instituto de Servicios de Salud del Estado de Aguascalientes (ISSEA), esta información se preprocesa, limpio y ordeno para conseguir los recursos que se categorizaron o discretizaron para su procesamiento y análisis.

En esta investigación, se diseñó, desarrollo y se puso a punto un mecanismo de clasificación compuesto por dos mecanismos, uno basado en la técnica de selección de variables llamada testores típicos y el otro mecanismo basado en la técnica de aprendizaje automático llamada redes neuronales artificiales.

Mediante la utilización de la técnica testores típicos, se consideraron dos clases: paciente con diabetes y pacientes sin diabetes, se obtuvieron todos los posibles subconjuntos de características que describen a un paciente, mediante el cálculo del peso informacional [1] se evalúa la relevancia de cada variable.

Seguido de esto se empleó la técnica de redes neuronales artificiales utilizando las variables obtenidas por los testores típicos y así se crea un clasificador para clasificar pacientes con diabetes en la población de Aguascalientes.

En parte la motivación de esta investigación es el poder tener un temprano y oportuno diagnóstico de esta patología, ya que la diabetes mellitus es la primera causa de muerte a nivel mundial [2] y la segunda causa de muerte tanto en mujeres como en

hombres en México [3], esta patología ha llegado al nivel de ser comparado con epidemias.

En la actualidad, se cree que en conjunto las variables: Tensión arterial, nivel de glucemia, el índice de masa corporal y el género del paciente son los factores que determinan la presencia de la diabetes mellitus en la población de Aguascalientes.

Un rasgo destacado de esta investigación es que se han considerado los insumos de información manejados en el sistema de salud más importante del Estado de Aguascalientes llamado Expediente Clínico Electrónico, basados en CIE10 [1].

## **2. Conceptos relacionados**

### **2.1. Diabetes mellitus**

La diabetes mellitus es una enfermedad crónica en la cual el cuerpo humano no utiliza bien la insulina producida por el páncreas o el mismo páncreas no produce insulina, la cual es una hormona que interviene como válvula permitiendo que la glucosa de los alimentos ingeridos pase de la sangre a las células para la producción de energía, la glucosa es asistida por la insulina para entrar a las células, es así como la falta de insulina se refleja en un exceso de glucosa [3].

Los bajos niveles de producción de insulina o de utilizarla de manera eficaz lleva a presentar hiperglucemia o niveles elevados de glucosa en la sangre. Manifiestar altos niveles de glucosa durante un tiempo prolongado se relacionan con casos de pacientes con daños corporales y fallas en algunos órganos y tejidos [2].

De acuerdo con la Federación Internacional de Diabetes, China, India, Estados Unidos, Brasil, Rusia y México, son en ese orden los países con mayor número de diabéticos [3]. La Organización Mundial de la Salud posiciona a la diabetes mellitus a nivel mundial como la Octava causa de muerte del año 2000 al 2012, el INEGI sitúa a la diabetes mellitus como una de las 3 causas de muerte desde 2006 a nivel nacional, el Gobierno del Estado de Aguascalientes la posiciona como la tercer causa en mortalidad general y en primer lugar como causa única de muerte [2].

En 2016 se realizó un estudio acerca niveles de diabetes en todo el mundo, dirigido por científicos del Imperial College de Londres, de la escuela de salud pública Harvard T.H. Chan, de la OMS y otros 500 investigadores de todo el mundo este estudio menciona que en 2016 el costo anual del tratamiento de diabetes fue de más de 825 billones de dólares nivel mundial, 170 billones en china, 105 billones en Estados Unidos de América y 73 billones de dólares en la India.

### **2.2. Testores típicos**

Cuando se cuenta con un grupo de datos con demasiadas clases se incrementa la probabilidad de que algoritmos de clasificación encuentren subgrupos de características que no son totalmente correctos [4, 5]. Los métodos de selección pueden acortar el tiempo de entrenamiento de un modelo clasificador [6], de esta manera se simplifica la interpretación a los usuarios [7], se evita el exceso de dimensión y se generaliza por reducción de varianza.

Los métodos de selección de variables se separan en filter, wrapper e híbridos. Los Métodos wrapper: estos se utilizan para medir o conducir el proceso de búsqueda de los diferentes subconjuntos de variables, utilizando un algoritmo de agrupamiento y evaluando los resultados mediante criterios específicos. Estos califican las variables en subconjuntos y no solo cada variable por separado.

Los testores típicos se encuentran ubicados en estos métodos de selección de variables wrapper ya que seleccionan los subgrupos de variables y su relación evaluando todas las variables del subgrupo, estos consideran el subconjunto y su interrelación multivariable.

El concepto de testor aparece a mediados de los años cincuenta, fue empleado en la detección de fallas de circuitos eléctricos por Cheguis y Yablonskii [8], además se utilizó en selección de variables en problemas de Geología [9]. La ponderación de importancia de cada una de las variables relacionadas a un problema puede ser alcanzado mediante el cálculo del peso informacional [10].

Considerando que  $U$  es un conjunto de objetos descritos por  $N$  características, agrupados en  $K$  clases. En base a la comparación de cada característica de los objetos que pertenecen a una clase contra los que pertenecen a las demás, se crea la matriz de diferencias MD. MD se construye por un criterio de comparación o de igualdad estricta de rasgos [11].

Una vez que se crea la MD, se genera la matriz básica MB, que está constituida por las filas de MD que son básicas, esto es:

Una fila  $i_q$  es básica si no existe alguna fila  $i_p$  que sea subfila de  $i_q$ .

Sean  $i_p$  e  $i_q$  filas de MD.

Se dice que  $i_p$  es una subfila de  $i_q$  si para todo elemento de  $i_q = 0$  se cumple que  $i_p = 0$  y además, existe por lo menos un elemento de  $i_q = 1$  en el que  $i_p = 0$  [12].

Una fila de la MB es un testor típico, si al eliminar cualquiera de sus características, pierde la propiedad de ser un testor, como citan Santiesteban y Pons en 2003 [13].

Un testor típico es un testor en su mínima expresión o aquel testor al que se le ha eliminado toda redundancia [14].

### 2.3. Redes neuronales artificiales

En la actualidad las técnicas de machine learning presentan incremento en su utilización tanto en el ámbito empresarial como el ámbito académico, ya que son métodos capaces de detectar automáticamente patrones en la información, ya sea para predecir o tomar decisiones.

Una red característica radica en una secuencia de capas con conexiones entre capas contiguas consecutivas, general mente las redes neuronales artificiales (RNA) están organizadas en capas, de las cuales dos capas son de contacto con el exterior, una capa de entrada en la que ingresan los datos a la red, y una capa de salida la cual entrega la respuesta de la red de entrada, el resto de las capas de la red son llamadas capas ocultas.

Una RNA puede ser o no densamente interconectada, y tiene los siguientes elementos, 1.-Entrada, la cual recibe la información desde el exterior, 2.- Pesos, cada entrada tiene su propio peso relativo o importancia de la entrada, 3.-Suma ponderada y 4.-Salida.

Para desarrollar una RNA no hay que programar ni el conocimiento ni las reglas del procesamiento del conocimiento, la RNA aprende las reglas del procesamiento del

conocimiento a través de ajustes de conexiones ponderadas entre las neuronas de las distintas capas de la red.

Existen dos fases en la modelización de RNA como lo menciona Marín Diazaraque [15]. Entrenamiento: se utiliza un conjunto de datos de entrenamiento para determinar los pesos que definen el modelo de red neuronal.

Prueba: es la fase donde se evalúa la precisión del modelo entrenado y para esta fase se utiliza un grupo de datos distintos a los de entrenamiento, ya que puede haber sobre ajuste que es cuando el modelo se ajusta demasiado a las singularidades en los patrones de entrenamiento y por lo cual pierde la habilidad de generalizar su aprendizaje a casos nuevos.

Las neuronas artificiales tienen diferentes estados de activación, comúnmente son dos, pero otras pueden tomar un valor de un conjunto definido. Con esta función de activación se evita que el resultado de una red neuronal sea idéntico a solo tener una sola neurona, para esto se necesita que cada neurona pase por una transformación no lineal. Estas transformaciones pueden ser: lineal, relu o unidad lineal rectificadora, sigmoide o tangente.

## **2.4. CIE 10**

Clasificación Estadística Internacional de Enfermedades y Problemas Relacionados con la Salud Décima Revisión o CIE 10, establece un registro sistemático, interpretación, análisis y comparación de los datos de la mortalidad y morbilidad en diferentes países y épocas. Permite clasificar patologías y problemas relacionados con la salud, así como diagnósticos, razones de admisión, afecciones tratadas, motivo de la consulta e incluso las causas de defunción [16].

## **3. Trabajos relacionados**

### **3.1. Reglas de predicción clínica**

Tradicionalmente una gran cantidad de médicos tratan a la diabetes Mellitus y otras patologías mediante las Reglas de Predicción Clínica (RPC), las cuales son herramientas de diagnóstico, pronóstico o intervención que incluyen las características clínicas ponderadas del historial clínico, examen físico y exámenes de gabinete [17], es decir que se basan en la forma en que se han diagnosticado a otros pacientes y el modo en el que han evolucionado.

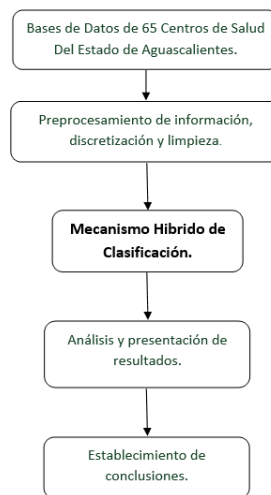
Esto se hace con base en los conocimientos de casos del médico o mediante casos que han sido publicados, pues la información sobre esta enfermedad está en constante cambio, médicos colaboran en descubrimientos y estudios, información que es relevante, aunque, esto hace que sea complicado tomar en cuenta todos los datos con los que se encuentran disponibles.

Debido a que las reglas de predicción clínica no se encuentran sistematizadas en una herramienta que asista al médico durante la consulta, el desarrollo de un mecanismo clasificador mediante testores típicos y redes neuronales artificiales puede ser una herramienta útil para mejorar la forma en que se diagnostica esta enfermedad.

## 4. Material y métodos

### 4.1. Metodología

Esta investigación conto con la metodología representada por medio de la siguiente Grafica en la Fig.1:



**Fig. 1.** Etapas de la metodología utilizada.

En esta metodología se hicieron los siguientes pasos:

1. Se extrajo la información de consultas de cada uno de los 65 centros de salud urbanos y rurales pertenecientes al ISSEA, la cual se concentró en una sola base de datos.
2. Se preprocesa la información reunida en la base de datos, eliminando registros incompletos, con error de captura además se eliminaron variables con datos fuera de rango, nulos y con datos sensibles de los pacientes. Se identificaron los pacientes con diagnostico confirmado de diabetes mellitus y los pacientes sin diagnóstico de diabetes y sin diagnósticos de alteraciones en glucosa.
3. Se procesa la información por medio del Mecanismo de Clasificación.
4. Se analizaron y presentaron resultados.
5. Se establecieron conclusiones.

### 4.2. Preparación de la información

Se comenzó filtrando la información de todos pacientes, dejando los confirmados diagnosticados con diabetes mellitus y sin diagnóstico de diabetes mellitus su historial y/o enfermedades relacionadas con alteraciones de glucosa.

Se contó inicialmente con 18 variables de signos vitales y métricas de la consulta, identificación geográfica y de derechohabencia del paciente y consulta, de las cuales se consideraron solo 12 variables ya que se excluyeron consultas con datos incompletos, nulos y variables con datos sensibles.

La información de consultas obtenida de los centros de salud se concentró en una única base con datos, reuniendo más de 100,000 consultas de todo tipo de patologías, de las cuales se extrajo una muestra de 500 pacientes confirmados con diabetes mellitus y otra de 500 pacientes sin diagnóstico de diabetes mellitus en su historial ni diagnósticos relacionados con alteraciones de glucosa.

### **4.3. Discretización de la información**

Las 12 variables consideradas para esta investigación se enlistan a continuación.

**Derechohabencia:** esta variable señala la institución de pertenencia de prestaciones de servicios de salud a la que el paciente es derechohabiente esta variable está categorizada por el Instituto Nacional de Estadística y Geografía (INEGI) de la siguiente manera: 0.- Población Abierta, 1.-Seguro Popular, 2.-IMSS, 3.-ISSSTE, 4.-PEMEX, 5.-SECMAR, 6.-Otra y 7.-SMNG. Este dato es de índole socio económica ya que algunos centros de salud tienen tendencias marcadas hacia alguna derechohabencia dependiendo de su ubicación.

**Sexo:** esta variable representa el género del paciente y está definida como 1.- Masculino y 2.- Femenino.

**Municipio de residencia:** es el número de municipio catalogado por el INEGI de la siguiente manera, 1.- Aguascalientes 2.-Asientos 3.-Calvillo, 4.-Cosío, 5.-Jesús María, 6.-Pabellón de Arteaga, 7.-Rincón de Romos, 8.-San José de Gracia, 9.-Tepezala, 10.-El Llano, 11.-San Francisco de los Romo. A través de esta variable se puede saber la incidencia poblacional municipal, así como el distanciamiento con respecto a la capital del estado.

**Jurisdicción Sanitaria:** es una clasificación de las tres jurisdicciones existentes en el estado de Aguascalientes, 1.- Jurisdicción I, 2.- Jurisdicción II y 3.- Jurisdicción III. Con esta variable se agrupan los municipios como tres únicos grupos, así como medio de análisis presupuestal ya que cada uno cuenta con su propio presupuesto anual.

**Existencia de antecedentes familiares con diabetes:** es una variable booleana la cual señala con 1 cuando el paciente tiene antecedentes y con 0 cuando el paciente no tiene antecedentes.

**Nivel de glucemia:** es el nivel de riesgo en salud del paciente determinado por la cifra de azúcar o glucosa en una muestra de sangre. el ISSEA la clasifica de la siguiente manera, 1.-Sin Riesgo, 2.- Normal en control, 3.- Con Riesgo.

**Estado Civil:** esta variable es la categoría de un individuo según el registro civil en función de si tiene o no pareja o su situación legal al respecto. El INEGI la cataloga de la siguiente manera, 1.-Soltero, 2.-Casado, 3.-Unión Libre, 4.-Separado, 5.-Divorciado, 6.-Viudo, 7.-Desconocido.

**Índice de Masa Corporal: (IMC)** es un indicador simple de la relación entre el peso y la talla de un paciente, se utiliza frecuentemente para identificar si se tiene sobrepeso u obesidad. Se calcula dividiendo el peso de una persona en kilos por el cuadrado de su talla en metros ( $\text{kg}/\text{m}^2$ ). Se utilizó la clasificación establecida por el ISSEA: 1.- Desnutrido, 2.-Normal, 3.-Sobre Peso, 4.- Obesidad.

Complicación: esta variable confirma si el paciente ha tenido alguna complicación en general durante su historial clínico y se describe con 1 con complicación y 0 sin complicación.

Edad: Es la edad del paciente en años. Para esta variable se consideró la clasificación propuesta por Mansilla. A [21] 1.-joven (< 24 años), 2.- Adulto (>= 24 años < 65 años), 3.- Senecto (>= 65 años < 75 años), 4.-Anciano (>= 75 años).

Tensión arterial baja y tensión arterial baja: Casiano Parra ordena los niveles de tensión arterial en el ser humano de la siguiente: 1.-Tensión Óptima, 2.-Tensión Normal, 3.- Tensión Normal Alta, 4.-Hipertensión Grado 1, 5.-Hipertensión Grado 2, 6.-Hipertensión Grado 3, 7.- Hipertensión Grado 4 [22].

En la Tabla 1, se listan las variables utilizadas en la investigación.

**Tabla 1.** Variables utilizadas.

Variable	Nivel	Dominio	Criterio de comparación
Derechohabiciencia	Poblacional	0,1,2,3,4,5,6,7	Igualdad Estricta
Sexo	Poblacional	1,2	Igualdad Estricta
Municipio	Poblacional	1,2,3,4,5,6,7,8,9,10,11	Igualdad Estricta
Jurisdicción	Poblacional	1,2,3	Igualdad Estricta
Ant. Fam. Diab.	Clínico	0,1	Igualdad Estricta
Nivel Glucemia	Clínico	1,2,3,4,5,6	Igualdad Estricta
Estado Civil	Poblacional	1,2,3,4,5,6,7	Igualdad Estricta
IMC	Clínico	1,2,3,4	Igualdad Estricta
Complicación	Clínico	0,1	Igualdad Estricta
Edad	Poblacional	1,2,3,4	Igualdad Estricta
Tensión A. Baja	Clínico	1,2,3,4,5,6,7	Igualdad Estricta
Tensión A. Alta	Clínico	1,2,3,4,5,6,7	Igualdad Estricta

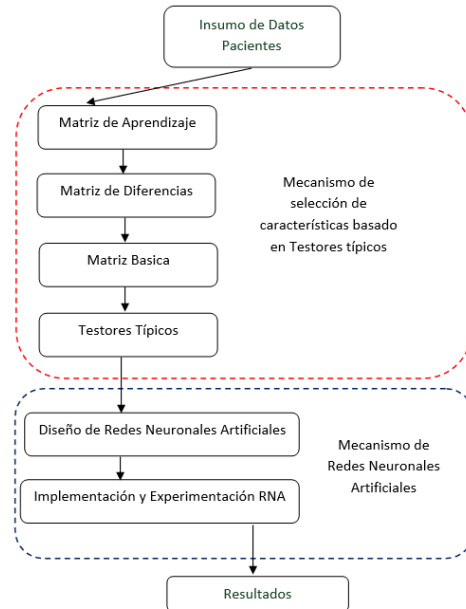
#### 4.4. Mecanismo de clasificación

Se desarrolló un mecanismo de clasificación compuesto por dos mecanismos secuenciales, el primer mecanismo fue testores típicos para selección subconjuntos de características, el segundo mecanismo integrado fue un clasificador basado en redes neuronales artificiales.

## 5. Resultados

Como parte del resultado se diseñó, desarrollo e implemento el siguiente marco de trabajo para el mecanismo de clasificación ilustrado en la Fig. 2.





**Fig. 2.** Etapas de marco de trabajo.

El mecanismo de clasificación está integrado por dos mecanismos:

El primer mecanismo fue testores típicos el cual selecciona subconjuntos de características mediante los siguientes pasos:

1. Se obtuvo la Matriz de aprendizaje a partir de los datos discretizados de las consultas de pacientes de las dos clases establecidas.
2. Se generó la Matriz de diferencias.
3. Se generó la Matriz Basica.
4. A partir de esa matriz básica se obtuvieron todos los testores y se identificó solo a los testores típicos.

Los testores típicos poseen como ventaja la dependencia total de los datos, no tienen límite de clases y de características a analizar, no solo determinan las variables importantes estos además determinan el peso informacional de cada una de las variables de importancia.

Se implementó una selección de características haciendo uso de los testores típicos asociados con la matriz de entrenamiento extraída de un archivo histórico de pacientes con la patología de diabetes, de pacientes que no la padecen de los cuales se analizaron un total de 12 variables que pueden ser consultadas en la Tabla 2.

Se utilizaron 500 registros de la Clase Diabéticos (1) y 500 registros de la clase No Diabéticos (0).

Los datos se procesaron mediante la técnica de testores típicos la cual fue desarrollada en RadStudio 10.4 en el lenguaje C++, y se obtuvieron los siguientes 5

**Tabla 2.** Variables iniciales para testores.

Todas las Variables	
VARIABLES	DATO
1	Derechohabiencia
2	Sexo
3	Municipio
4	Jurisdicción
5	Ant. Fam. Diab
6	Nivel Glucemia
7	Estado Civil
8	IMC
9	Complicación
10	Edad
11	Tensión A. Baja
12	Tensión A. Alta

**Tabla 3.** Testores típicos obtenidos.

Testor 1	Sexo	Glucemia	Edo. civil	Imc	Comp.	Tensión Alta
Peso inf.	40%	80%	100%	60%	100%	40%
Testor 2	Glucemia	Estado civil	Imc	Comp.		
Peso inf.	80%	100%	60%	100%		
Testor 3	Ant. Fam. Diab.	Estado civil	Comp.	Tensión Alta		
Peso inf.	40%	100%	100%	40%		
Testor 4	Sexo	Glucemia	Edo civil	Imc	Comp.	
Peso inf.	40%	80%	100%	60%	100%	
Testor 5	Ant. Fam. Diab.	Glucemia	Edo civil	Comp.	Edad	
Peso inf.	40%	80%	100%	100%	40%	

testores típicos o subgrupos de características que se muestran en las siguiente Tabla 3. En éstas se presentan las variables que integran el testor y su peso informacional.

Para el mecanismo de clasificación de pacientes diabéticos se utilizaron redes neuronales artificiales ya que son clasificadores superiores o excelentes reconocedores de patrones, estas son usadas donde no funcionan las técnicas tradicionales, manejan un amplio rango de datos, exclusiones y entradas deficientes. Se diseñó e implementó una red neuronal artificial densamente interconectada con una sola capa oculta para procesar la información de cada uno de los testores típicos, tomado como neuronas de

entrada cada una de las variables, implementando la ecuación 1 para calcular las neuronas de la capa oculta la cual propuesta por Piedra Fernández [18] , este autor propone 2 ecuaciones para capas ocultas, una para una sola capa y otra para 2 capas ocultas. Como neurona de salida la predicción de la clase diabético o no diabético:

$$h = n - m , \tag{1}$$

donde:  $h$  = número de neuronas de la capa oculta,  $n$  = número de neuronas de entrada, y  $m$  = número de neuronas de salida.

Las redes neuronales se desarrollaron en Python 3.0 utilizando las librerías Keras y Tensorflow. Los parámetros utilizados para el proceso de entrenamiento de las redes neuronales se utilizaron 100 registros de la clase diabéticos y 100 registros de la clase no diabéticos, aplicando activación relu tanto para las neuronas en la capa de entrada como en la capa oculta y activación sigmoide en la capa de salida, en el ajuste se procesó por 80 épocas y lotes de 15, para la fase de validación se utilizaron 30 registros de pacientes diabéticos y 30 registros de pacientes no diabéticos, dado que la selección de variables conlleva un proceso de aprendizaje con el total de los elementos de pacientes con y sin diabetes; las cantidades de registros de pacientes utilizadas para los procesos de aprendizaje y validación del clasificador fueron menores. La configuración topológica y los resultados de precisión de test de cada una de las redes neuronales se presentan en la Tabla 4.

**Tabla 4.** Configuración de redes neuronales artificiales para testores típicos.

Testor	Variables	Nodos Entrada	Nodos Oculta	Nodos Salida	Precisión de Test
1	6	6	5	1	0.5
2	4	4	3	1	0.883
3	5	5	4	1	0.883
4	5	5	4	1	0.883
5	5	5	4	1	0.916

Analizando los resultados de los 5 clasificadores desarrollados, el mejor resultado es el observado por la red neuronal artificial procesando las variables del Testor 5, el cual obtiene casi el 92% de precisión del test, determinando a las variables Tabla 5 *antecedentes familiares con diabetes, nivel de glucemia, estado civil, existencia de complicación y la edad*, en conjunto como las más importantes para la clasificación de diabetes en la población analizada, las cuales fueron analizadas.

Una de las variables con mayor peso informacional es la de estado civil presenta información relevante ya que se encontró que el 48% de los pacientes diabéticos son de la categoría *casados* y que solo el 22% de diabéticos son de la categoría *solteros*, el 30% restante pertenece a las otras clases restantes de estado civil. La otra variable con el mayor peso informacional fue la de existencia de complicación, en esta variable se muestra que el 90% de los pacientes con diabetes pertenecen a la clase *con complicación* y solo el 10% de los pacientes diabéticos pertenecen a la clase *sin*

*complicación*. Respecto al nivel de glucemia se pudo notar que el solo el 17% de la población con diabetes corresponde a la categoría *sin riesgo* y el 83% de la población con diabetes presenta categoría *con riesgo*.

En la variable edad los casos de pacientes diabéticos se centran la mayoría en las clases *adulto* y *senecto* que va de los rangos de los  $\geq 24$  años a los  $< 75$  años, estos se dividen en 61% de diabéticos en la clase *adulto* y 28% en la clase *senecto*, el otro 11% pertenecen al resto de las clases de edad.

**Tabla 5.** Peso Informacional de Testor de mejor precisión.

Variable	Descripción	Peso Inf.
1	Antecedentes familiares con diabetes	40%
2	Nivel de glucemia	80%
3	Estado civil	100%
4	Existencia de complicación	100%
5	Edad	40%

Por otro lado, la variable con antecedentes familiares con diabetes presenta un 67% de pacientes con diabetes de la clase *con antecedentes familiares* y un 33% de los pacientes con diabetes son de la clase *sin antecedentes*.

## 6. Conclusiones

Se encontró con éxito el subconjunto de características ponderadas que mejor describen a un paciente con diabetes mellitus en la población de Aguascalientes clasificado con diagnósticos del CIE10, se diseñó e implementó creó y se puso a punto un clasificador basado en redes neuronales artificiales con una precisión del 91.6%.

Se puede concluir que no se debe prescindir de las variables de estado civil y existencia de complicación ya que fueron las de mayor importancia (100% de peso informacional), además de reafirmar la importancia de las variables nivel de glucemia, antecedentes de familiar con diabetes y la edad, ya que en conjunto las 5 variables del testor resultante considera las relaciones entre todas las variables para identificar pacientes con diabetes mellitus.

Se descubrió que el estado civil tiene una relación muy directa en la descripción de un paciente con diabetes, y queda comprobado que si es una variable importante ya que cuenta con un peso informacional de 100% y la cual en su categoría casado tiene una alta incidencia de pacientes con diabetes y mucho menor incidencia en la categoría soltero, posiblemente la alta incidencia es detonada por los altos niveles de estrés y el marcado sedentarismo que presentan los pacientes de la categoría casado en la población de Aguascalientes.

Se concluye que las variables: historial de familiar con diabetes y la existencia de complicación tienen una alta relación, ya que cuando estas se encuentran en su categoría

con familiar y con complicación, de un total de 65 pacientes 62 fueron diagnosticados con la patología de diabetes. Cuando estas variables se presentan en sus categorías sin familiar y sin complicación de un total de 45 pacientes solo 5 fueron diagnosticados con diabetes.

Se rechaza la creencia respecto a que en conjunto las variables: Tensión arterial, nivel de glucemia, el índice de masa corporal y el género del paciente son los factores determinantes de diabetes mellitus en la población de Aguascalientes.

Para trabajos futuros se está considerando el integrar este tipo de clasificadores como recomendadores auxiliares asistiendo a médicos en el sistema de consultas de especialidad en el módulo de enfermedades crónico-degenerativas, el cual es uno de los módulos de mayor uso del Expediente Clínico Electrónico del Instituto de Servicios de Salud.

## Referencias

1. OPS: Clasificación internacional de las enfermedades. Organización panamericana de la salud. [http://www.paho.org/hq/index.php?option=com\\_content&view=article&id=3561%253Aclasificacion-internacional-enfermedades-cie&catid=2641%253Acha-clasificacion-internacional-enfermedades-cie&Itemid=2560&lang=en](http://www.paho.org/hq/index.php?option=com_content&view=article&id=3561%253Aclasificacion-internacional-enfermedades-cie&catid=2641%253Acha-clasificacion-internacional-enfermedades-cie&Itemid=2560&lang=en) (2015)
2. INEGI: Estadísticas a propósito del día mundial de la diabetes (2011)
3. Herrera, F.: Introducción a los Algoritmos Metaheurísticos. *Soft Comput. Intell. Inf. Syst.* (2009)
4. Jensen, R., Shen, Q.: *Computational intelligence and feature selection : rough and fuzzy approaches* (2008)
5. Liu, H., Motoda, H.: *Computational methods of feature selection* (2008)
6. Bermingham, M.L., Pong-Wong, R., Spiliopoulou, A., Hayward, C., Rudan, I., Campbell, H., Wright, A.F., Wilson, J.F., Agakov, F., Navarro, P., Haley, C.S.: Application of high-dimensional feature selection: evaluation for genomic prediction in man. *Sci. Rep.*, 5, pp. 1–12 (2015)
7. Casella, G., Fienberg, S., Olkin, I.: *An Introduction to Statistical Learning* (2013)
8. Cheguis, I.A., Yablonskii, S.V.: About testors for electrical outlines. *Uspeji Mat. Nauk*, 4, pp. 188–184 (1955)
9. Alba, C., Santana, R., Ochoa, R., Lazo, C.M.: Finding typical testors by using an evolutionary strategy. In: *Proc. V Iberoam. Work. Pattern Recognit.*, pp. 267–278 (2000)
10. Torres, M., Ponce, E., Ochoa, C., Torres, A., Díaz, E.: Mecanismos de aceleración en selección de subconjuntos de características basada en el peso informacional de las variables para aprendizaje no supervisado (2009)
11. Soto-Torres, M.D., Soto-Torres, A., Sentí, E.P.L.: Algoritmo genético y testores típicos en el problema de selección de subconjuntos de características. *Sist. Cibernética E Informática*, 3(2), pp. 1–5 (2006)
12. Shulcloper, J.R., Alba, C., Lazo, C.: *Introducción a la teoría de testores típicos* (1995)
13. Santiesteban, A., Pons, P., Lex, A.: Un nuevo algoritmo para el cálculo de los testores típicos. *Revista Ciencias Matemáticas V*, 21, pp. 88–95 (2003)
14. Torres, M.D.: *Metaheurísticas híbridas en selección de subconjuntos de características para aprendizaje no supervisado*. Universidad Autónoma de Aguascalientes (2010)
15. Marín-Diazaraque, J.M.: *Introducción a las redes neuronales aplicadas Conceptos básicos*, 3 (2003)
16. Vélez, D.: *Estructura y uso del CIE10* (2016)
17. Karzulovic, L. García, P., Wozniak, A., Villarroel, L., Hirsch, T., Concha, I., Catalán, S., Cifuentes, L.: Una regla de predicción clínica anticipa el diagnóstico de la faringitis

*Pablo Rodríguez de León, María Dolores Torres Soto, Aurora Torres Soto*

- estreptocócica en niños de 2 a 15 años?. Rev. Chil. infectología, 35(5), pp. 476–482 (2018)
18. Piedra Fernández, J.A.: Aplicación de los sistemas neurodifusos a la interpretación automática de imágenes de satélite. Universidad de Almería (2008)