



Implementación de mapas cognitivos difusos con algoritmos genéticos para predecir diabetes mellitus tipo 2

Implementation of Fuzzy Cognitive Maps with Genetic Algorithms for Predicting Type 2 Diabetes Mellitus

  William Hoyos¹;
 Rander Ruíz²;
 Kenia Hoyos³

¹Universidad Cooperativa de Colombia, Montería-Colombia,
william.hoyos@campusucc.edu.co

²Universidad de Antioquia, Cauca-Colombia,
rander.ruiz@udea.edu.co

³Laboratorio Clínico Humano, Clínica Salud Social,
Sincelejo-Colombia,
kmhoyosgonzalez@gmail.com

Cómo citar / How to cite

W. Hoyos, R. Ruíz, and K. Hoyos, "Implementación de mapas cognitivos difusos con algoritmos genéticos para predecir diabetes mellitus tipo 2," *Tecnológicas*, vol. 27, no. 60, e3061, 2024.
<https://doi.org/10.22430/22565337.3061>

Resumen

La diabetes mellitus tipo 2 es una enfermedad crónica no transmisible, causada por un trastorno en el metabolismo de la glucosa, que provoca un aumento anormal de su concentración en la sangre. El diagnóstico tardío de esta enfermedad contribuye al aumento de las tasas de morbilidad y mortalidad a nivel mundial. El desarrollo de modelos basados en inteligencia artificial para la predicción de diabetes podría acelerar el diagnóstico. Por tanto, el objetivo del presente estudio fue implementar un modelo de predicción de diabetes mellitus tipo 2 basado en mapas cognitivos difusos entrenado con un algoritmo genético. La metodología empleada consistió en utilizar un conjunto de datos del Instituto Nacional de Diabetes y Enfermedades Digestivas y Renales de la población de indios PIMA, que contiene información demográfica y clínica de 768 pacientes. El 70 % de los datos se empleó para el entrenamiento y validación, y el 30 % restante se utilizó para las pruebas de rendimiento. El modelo de mapas cognitivos difusos puede predecir la enfermedad con un 99 % de exactitud, 98 % de precisión y recall de 100 %. Se concluye que el modelo presenta una buena capacidad para predecir y evaluar el comportamiento de las variables de interés en la diabetes mellitus tipo 2, mostrando su valor como herramienta de soporte en la identificación oportuna de la enfermedad y apoyo a la toma de decisiones por parte del profesional médico.

Palabras clave

Diabetes Mellitus Tipo 2, mapas cognitivos difusos, factores de riesgo, algoritmos de predicción, algoritmos genéticos.

Abstract

Type 2 diabetes mellitus is a chronic non-communicable disease caused by a disorder in glucose metabolism, which results in an abnormal increase in glucose concentration in the blood. The late diagnosis of this disease contributes to the increased worldwide rates of morbidity and mortality. The development of models based on artificial intelligence for the prediction of diabetes could accelerate diagnosis. Therefore, the aim of the present study was to develop a prediction model for type 2 diabetes mellitus based on fuzzy cognitive maps trained with a genetic algorithm. The methodology employed consisted of using a dataset from the National Institute of Diabetes and Digestive and Kidney Diseases PIMA Indian population, which contains demographic and clinical information from 768 patients. For training and validation, 70 % of the data was used and the remaining 30 % was used for performance testing. The fuzzy cognitive map model can predict the disease with 99 % accuracy, 98 % precision, and 100 % recall. It is concluded that the model presents a good ability to predict and evaluate the behavior of the variables of interest in type 2 diabetes mellitus, showing its value as a support tool for the timely identification of the disease and support for decision making by the physician.

Keywords

Type 2 Diabetes Mellitus, fuzzy cognitive maps, risk factors, prediction algorithms, genetic algorithms.

1. INTRODUCCIÓN

La diabetes mellitus tipo 2 (DM2) es un trastorno metabólico crónico caracterizado por presentar resistencia a la insulina y una producción anormal de esta hormona, lo que provoca niveles elevados de glucosa en sangre [1]. En los últimos años, la diabetes ha incrementado su prevalencia convirtiéndose en un problema de salud pública a nivel global [2]. Se calcula que en 2021 había 537 millones de personas con diabetes en el mundo, de las cuales más del 95 % corresponde al tipo 2, produciendo alrededor de 6,7 millones de muertes [3].

El principal problema asociado a la DM2 es su significativa repercusión tanto en la salud individual como en los sistemas de salud pública, en muchos casos debido a la baja oportunidad para el diagnóstico y manejo clínico. Por lo tanto, un diagnóstico tardío puede generar complicaciones que llevan a la muerte [2]. Las personas con DM2 enfrentan un mayor riesgo de desarrollar diversas complicaciones como la enfermedad cardiovascular, neuropatía, enfermedad renal, retinopatía y enfermedad vascular periférica, lo que desmejora significativamente la calidad de vida con aumento de la tasa de mortalidad entre estos individuos [1]. A pesar de los avances en las modalidades de tratamiento, lograr y mantener un control glucémico óptimo en muchos pacientes con DM2 puede ser todo un desafío. La adherencia a los regímenes de medicación, las modificaciones en el estilo de vida y el monitoreo regular son cruciales para un manejo exitoso, pero pueden ser difíciles de lograr en algunos entornos del mundo real [3]. En consecuencia, la diabetes impone una carga económica considerable a los sistemas de salud a nivel mundial, pues los costos médicos directos asociados con el diagnóstico y manejo de la diabetes y sus complicaciones, así como los costos indirectos relacionados con la pérdida de productividad y la discapacidad, son significativos [4], [5].

Para abordar esta problemática, se han publicado diversos estudios para la predicción y clasificación de la diabetes [6]. En [7] desarrollaron un sistema de monitoreo inteligente empleando una metodología de aprendizaje de conjunto (ensemble learning) con algoritmos genéticos (AG), que consistía en combinar varios algoritmos de aprendizaje automático (ML, por sus siglas en inglés) para mejorar las predicciones. Los resultados demostraron que el modelo propuesto supera significativamente a los enfoques de ML individuales y a los enfoques de conjunto convencionales, con una exactitud de 98 %. En [8] desarrollaron modelos de bosque aleatorio (RF, por sus siglas en inglés) y regresión logística (LR, por sus siglas en inglés) para la predicción de DM2, usando los conjuntos de datos de Sylhet (contiene información de pacientes del Hospital de Diabetes de Sylhet, en Bangladesh) y PIMA (procede del Instituto Nacional de Diabetes y Enfermedades Digestivas y Renales, con información de pacientes de la etnia indígena Pima). Para el conjunto de datos PIMA el mejor resultado lo obtuvo RF con una exactitud de 78 %. En [9] evaluaron cuatro algoritmos de aprendizaje automático: LR, árbol de decisión (DT, por sus siglas en inglés), RF, y refuerzo de gradiente extremo (EXB, por sus siglas en inglés) para predecir complicaciones de diabetes. Los valores del área bajo la curva oscilaron entre el 60 % (retinopatía) y el 69 % (insuficiencia cardíaca congestiva), mientras que las tasas de exactitud variaron entre el 60 % (retinopatía) y el 75 % (hipertensión). La técnica más relevante para la predicción de hipertensión, el miocardio y la retinopatía fue RF.

Otros estudios han utilizado un enfoque no estadístico para estudiar el problema de la predicción de diabetes, como los mapas cognitivos difusos (FCM, por sus siglas en inglés) [10]. Por ejemplo, en [11] desarrollaron un modelo jerárquico de FCM como sistema de apoyo a la decisión médica para clasificar entre los tipos de diabetes. El enfoque se diseñó en 3 fases y obtuvo la mejor exactitud (83,3 %) al evaluar la diabetes tipo 1 y la DM2. En [12] se exploró la relación entre los factores de riesgo y los síntomas de la diabetes con el empleo de un FCM

basado en expertos. El estudio concluyó que la edad adulta, la obesidad y el tabaquismo son los factores que afectan en mayor medida la condición diabética. En [13] desarrollaron un modelo de inteligencia artificial (IA) basado en FCM para apoyar la toma de decisiones clínicas en la detección temprana de la diabetes. El modelo fue entrenado con el algoritmo de optimización por enjambre de partículas (PSO, por sus siglas en inglés), utilizando el conjunto de datos del hospital de diabéticos de Sylhet, Bangladesh. Los resultados demostraron un excelente desempeño predictivo, con una exactitud del 95 %. También permitió evaluar el comportamiento de las variables involucradas, lo que proporcionó información importante relacionada con la dinámica de los factores de riesgo asociados a diabetes.

Existen pocos reportes en la literatura relacionados con el desarrollo de modelos de FCM y algoritmos de optimización que motivaron la presente investigación. En primer lugar, [10]-[13] demostraron la utilidad de los FCM para la predicción de complicaciones de la diabetes, la toma de decisiones clínicas y el diseño de estrategias de intervención efectivas para mejorar los resultados de los pacientes. Sin embargo, [10]-[12] no analizaron el comportamiento de las variables que influyen en el desarrollo de la enfermedad. Tampoco evaluaron el desempeño predictivo de dichos modelos en conjuntos de datos reales, sino que se utilizaron datos simulados limitados.

En [13], la optimización con PSO involucró menos parámetros o variables susceptibles de ajuste, por lo que fue relativamente simple y de fácil implementación. No obstante, la rápida convergencia del algoritmo PSO durante las primeras etapas de búsqueda podría agotar su capacidad para explorar nuevas áreas del espacio de búsqueda de todas las posibles soluciones que pueden ser consideradas para un problema dado. En síntesis, PSO es más sencillo y rápido, pero puede tener problemas con la diversidad y la convergencia prematura [14]. Por otra parte, no hay suficientes estudios que relacionen el gran potencial de las optimizaciones de FCM con AG para el pronóstico de diabetes.

Por lo tanto, el objetivo de la investigación fue implementar un modelo de FCM optimizado con un AG para la predicción inteligente de DM2. El AG es un algoritmo más robusto que puede manejar una amplia variedad de problemas complejos, gracias a la diversidad genética en la población, lo que ayuda a evitar la convergencia prematura. Los AG exploran mejor el espacio de búsqueda de soluciones, desde luego, a costa de una mayor complejidad y tiempo de ejecución [14]. El enfoque propuesto logró una exactitud y precisión excelente, y tiene la posibilidad de funcionar como sistema de apoyo a la toma de decisiones médicas debido a su fácil interpretación y análisis. El desarrollo de un modelo con alta precisión y confiabilidad puede contribuir con la identificación temprana del riesgo de desarrollar la enfermedad. Además, se realizó una comparación cuantitativa del rendimiento del FCM con otros algoritmos de aprendizaje automático reportados en la literatura. Finalmente, el enfoque propuesto permite analizar la evolución de las variables a lo largo del tiempo, usando gráficas de iteraciones simuladas. El uso de estas herramientas en el campo de la salud podría tener un impacto positivo en la optimización de los recursos, reducción de los costos de los sistemas de salud, disminución de tasas de morbilidad de DM2, atención personalizada y mejora en la calidad de vida de los pacientes.

El documento está organizado de la siguiente manera: en la Sección 2, se detalla la metodología empleada en el estudio, describiendo el conjunto de datos, las técnicas de preprocesamiento, construcción del modelo basado en FCM y AG, configuración experimental y evaluación. En la Sección 3 se muestran los resultados experimentales y se comparan con trabajos relacionados. Finalmente, en la Sección 4 se presentan las conclusiones del estudio, limitaciones y trabajos futuros.

2. METODOLOGÍA

En esta sección, se describe el marco metodológico utilizado para la implementación de los FCM con AG. En la Figura 1 se puede observar la metodología general que incluye varias etapas, desde la recolección de los datos, la fase de preprocesamiento, el entrenamiento del modelo de FCM con un AG y su posterior evaluación en un conjunto de pruebas.

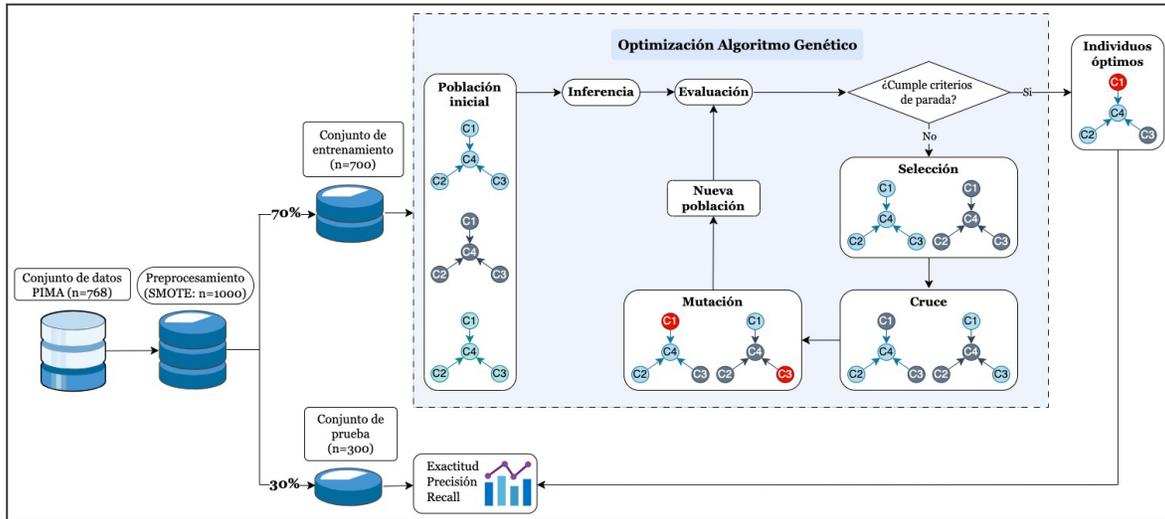


Figura 1. Metodología general para el desarrollo del FCM optimizado con AG. Fuente: elaboración propia.

2.1 Conjunto de datos

Se utilizó la base de datos de DM2 del Instituto Nacional de Diabetes y Enfermedades Digestivas y Renales de la población de indios Pima [15]. La información corresponde a 768 pacientes (500 sin DM2 y 268 con DM2) y 9 variables. En la Tabla 1 se observa la información de las 9 variables de estudio (incluida una variable de clase, que corresponde al diagnóstico de diabetes). Este conjunto de datos está disponible para su descarga desde internet [15], [16].

Tabla 1. Variables incluidas en el conjunto de datos PIMA. Fuente: elaboración propia.

Concepto	Nombre de la variable	Breve descripción de la variable
C1	Número de embarazos	Número de veces que la mujer se ha encontrado en estado de embarazo.
C2	Glucosa plasmática (TTOG)	Concentración de glucosa plasmática a las 2 horas, en un test de tolerancia oral a la glucosa (TTOG).
C3	Presión sanguínea diastólica (mmHg)	Presión entre latidos cuando el corazón se está llenando de sangre.
C4	Espesor del pliegue cutáneo del tríceps (mm)	Espesor de la piel que permite estimar la cantidad de tejido adiposo subcutáneo.
C5	Insulina sérica (μU/ml)	Medición de insulina sérica a las 2 horas
C6	Índice de Masa Corporal (kg/m ²)	Es un índice que se calcula con base en el peso y la estatura de la persona.
C7	Función del pedigrí de diabetes	Una función que califica la probabilidad de padecer diabetes basado en los antecedentes familiares
C8	Edad	Tiempo que ha pasado desde el nacimiento de un individuo.
C9	Diagnóstico	Proceso en el que se identifica una enfermedad, afección o lesión por sus signos y/o síntomas.

2.2 Preprocesamiento de datos

Teniendo en cuenta que el conjunto de datos consiste en información clínica de 500 individuos sin DM2 y 268 con DM2, el problema del desbalance de clases fue abordado generando instancias sintéticas de los individuos con DM2, con el fin de equilibrar el conjunto de datos. Para ello, se empleó la técnica de Sobremuestreo Sintético de Clase Minoritaria (SMOTE, por sus siglas en inglés) [17]. Esto garantiza que el modelo no genere resultados sesgados hacia la clase sobrerrepresentada, ayuda a encontrar los límites de decisión entre las clases y mejora el desempeño predictivo.

2.3 Construcción del modelo

2.3.1 Mapas Cognitivos Difusos (FCM)

Los FCM son una herramienta de IA que utiliza los aspectos principales de la lógica difusa y las redes neuronales artificiales. En los últimos años, han demostrado su potencial para modelar sistemas de apoyo en la toma de decisiones, analizar sistemas complejos por medio del razonamiento causal y facilidad para la representación del conocimiento [18], [19]. La construcción de los modelos de FCM se puede realizar mediante la definición de conceptos y asignación de relaciones a partir de expertos o por la optimización de las relaciones a través de algoritmos de aprendizaje automático [20].

La Figura 2 ilustra los componentes principales que integran un FCM. Este consta en primer lugar de nodos o conceptos (C), que representan las variables incluidas en el conjunto de datos PIMA (Tabla 1) [15]. Además, se observan las relaciones o pesos (W) en forma de una flecha dirigida desde un concepto de origen a un concepto de destino. El número como subíndice en la W de cada flecha indica la dirección de esas influencias. Por ejemplo, W_{14} representa la relación entre el Concepto 1 (C_1) y el Concepto 4 (C_4). La fuerza de la relación depende del valor asignado entre -1 a +1. Un valor de -1 indica un efecto inhibitorio entre los conceptos, un valor de +1 un efecto excitador y un valor de 0 indica que no existe relación causal los conceptos.

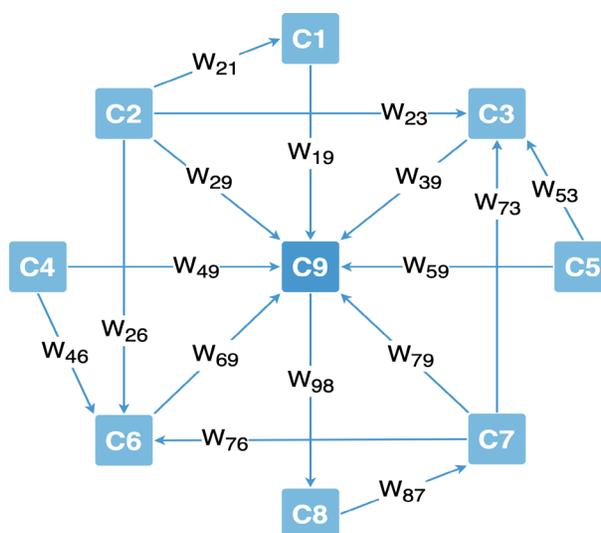


Figura 2. Estructura gráfica de un FCM simple, con nueve conceptos (nodos) y flechas que indican las relaciones entre ellos. Fuente: elaboración propia.

Matemáticamente, un FCM es definido mediante (1):

$$\Phi = \langle n, f(\cdot), r, W \rangle \quad (1)$$

donde n es el número de conceptos, W representa la matriz de pesos que define las relaciones entre los conceptos, $f(\cdot)$ es una función de activación que mantendrá los valores de los conceptos en un rango determinado por r .

El proceso de inferencia de los FCM consiste en la multiplicación sucesiva de un vector de activación $v(0)$ y la matriz W hasta que el sistema alcanza un estado estacionario. La ecuación 2 resume el proceso de inferencia o razonamiento del FCM:

$$a_j(t+1) = f\left(\sum_{i=1, i \neq j}^m W_{ij} a_i(t)\right) \quad (2)$$

donde $a_j(t+1)$ es el valor del concepto C_j en la iteración $t+1$, m es el número de conceptos, W_{ij} es el valor de la relación entre el concepto C_i y el concepto C_j , y $a_j(t)$ es el valor del concepto C_j en la iteración t .

2.3.2 Algoritmo genético (AG)

En el presente estudio, se empleó un AG para optimizar el modelo de FCM. Los AG son herramientas computacionales de búsqueda aleatoria basada en los principios de la genética y la selección natural biológica de Darwin. Es un algoritmo basado en población, que utiliza el concepto de supervivencia del más apto, en donde la alternativa óptima consta de los mejores componentes (genes) y se transmite de una generación a la siguiente [21]. Según el modelo, un cromosoma está representado por un código binario: 0 o 1. Un valor de 1 indica que la característica de referencia ha sido seleccionada, mientras que 0 indica lo opuesto [7]. El AG se aplica sobre una población de individuos (es decir, bits definidos), luego, los operadores de cruce y mutación generan nuevos candidatos en la población. En la última generación, el individuo más apto se considera la solución óptima.

La Figura 3 muestra la representación esquemática de las cinco etapas que se llevan a cabo durante la implementación de un AG. Dada una función de aptitud f , se busca maximizar (o minimizar) $f(x), x \in X$. Así, el proceso comienza con una población inicial aleatoria. Como ejemplo, tenemos 3 individuos con diferentes tipos de genes. En la etapa dos ocurre la evaluación usando la función de aptitud, lo que permite clasificar a la población como apta o no idónea. Tras el cálculo del valor de aptitud, continúa la etapa de selección para elegir a los mejores padres, los cuales siguen a las etapas de cruce y mutación. La operación de cruce se aplica para intercambiar información entre dos padres seleccionados. Durante la etapa mutación, se cambian los genes de los cromosomas de la descendencia cruzada.

Al finalizar las operaciones de selección, cruce y mutación, la nueva población se genera por completo y se obtendrán los individuos óptimos. El proceso es iterativo y se detendrá si hay convergencia de resultados o si el número de iteraciones supera el umbral máximo[22]-[24].

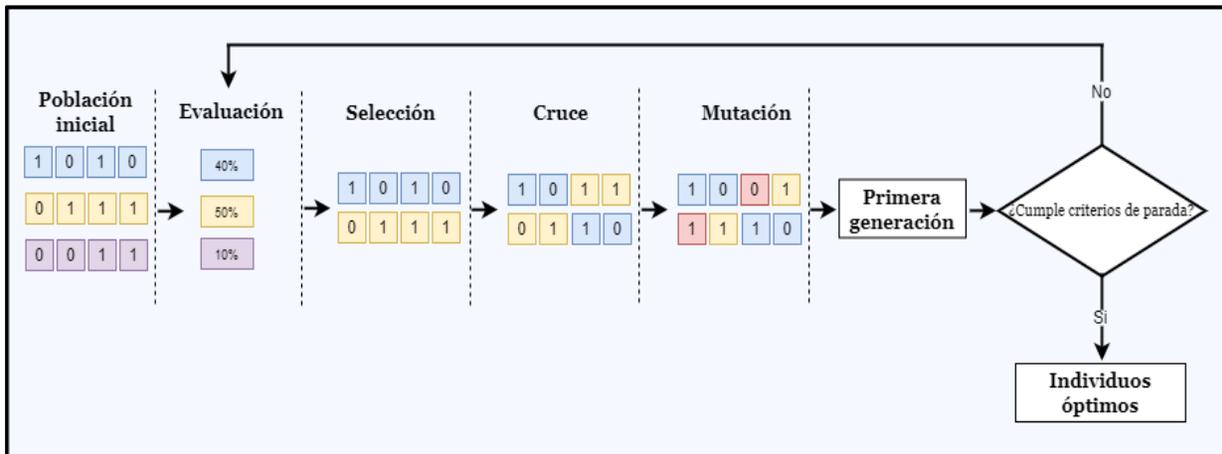


Figura 3. Representación gráfica de las etapas claves de un algoritmo genético. Fuente: elaboración propia.

2.3.3 Optimización del FCM con algoritmo genético

En la Figura 1, se observa la metodología empleada por el AG para la optimización del FCM. Inicialmente, se inicia una población aleatoria para asignar los pesos (W) del FCM teniendo en cuenta las variables del conjunto de datos. Los modelos de FCM obtenidos se someten a un proceso de inferencia utilizando tres componentes: la matriz de pesos o influencias ($W \in \mathbb{R}^{m \times m}$) que define la interacción entre los conceptos; una función de activación matemática que mantiene los valores de los conceptos en un rango; y un vector de estado ($a \in \mathbb{R}^m$). En esta etapa se calcula el vector de estado (a) a través de sucesivas multiplicaciones con la matriz de pesos, hasta que el sistema encuentre estabilidad. Tras el proceso de inferencia, los FCM son evaluados a través de métricas para medir el desempeño en el conjunto de datos de validación. Los FCM con mejor rendimiento son seleccionados para la etapa de cruce y mutación con el fin de desarrollar una nueva generación. El proceso se repetirá iterativamente hasta llegar a la condición de parada, que obtendrá el FCM con mejor desempeño.

2.3.4 Configuración experimental

El conjunto de datos preprocesados se dividió en un 70 % para el entrenamiento y validación del modelo. El 30 % restante de los datos, fue utilizado para la etapa de prueba del FCM seleccionado en la etapa de entrenamiento. La configuración de los hiperparámetros de las técnicas empleadas para la construcción de modelo, se pueden observar en la Tabla 2.

Tabla 2. Configuración de hiperparámetros utilizados para cada algoritmo. Fuente: elaboración propia.

Técnica	Hiperparámetro	Opciones de configuración
FCM	Función de activación	Sigmoide, tangente hiperbólica
	Función de inferencia	Kosko, Kosko modificado, Rescaled
	Población inicial	Valores aleatorios entre 20 y 200
AG	Probabilidad de cruce	Valores aleatorios entre 0 y 1
	Probabilidad de mutación	Valores aleatorios entre 0 y 1

2.4 Evaluación del modelo

2.4.1 Métricas de evaluación

Finalmente, el mejor modelo seleccionado se aplicó al conjunto de datos de prueba (30 %) para evaluar su rendimiento. El desempeño predictivo del modelo fue evaluado a través de métricas como exactitud, precisión, recall e índice Kappa. A continuación, se describen brevemente las métricas utilizadas para evaluar el rendimiento del modelo:

Exactitud: proporción de instancias que el modelo clasificó correctamente entre el número total de elementos. Una mayor exactitud significa un mayor rendimiento del modelo (3):

$$Exactitud = \frac{VP + VN}{VP + FN + FP + VN} \quad (3)$$

donde VP son los verdaderos positivos, VN los verdaderos negativos, FN los falsos negativos y FP los falsos positivos.

Precisión: proporción de verdaderos positivos entre la cantidad total de predicciones que el modelo ha clasificado como positivas (4).

$$Precisión = \frac{VP}{VP + FP} \quad (4)$$

Recall: proporción de verdaderos positivos clasificados correctamente entre el total de los casos que realmente son positivos (5).

$$Recall = \frac{VP}{VP + FN} \quad (5)$$

Índice Kappa: medida que compara los resultados de la clasificación con los valores asignados por azar. Tiene valores entre 0 y 1, entre más cercano a uno, más concordancia existe en la clasificación.

2.4.2 Simulación computacional basada en escenario

Por último, se realizaron iteraciones simuladas usando los paquetes *fcm* y *ggplot2* [25] para visualizar el comportamiento de las variables involucradas en la predicción. En este paso se utiliza un vector inicial (conjunto de variables o atributos de cada paciente) que se ingresa al modelo y tras el proceso de inferencia, se obtiene un vector final que indica el estado en que los signos y síntomas ya no evolucionan más [20]. Para asegurar la reproducibilidad de la presente investigación, todos los archivos y código usado en los experimentos están alojados en el siguiente [material suplementario](#).

3. RESULTADOS Y DISCUSIÓN

En esta sección, se presentan los resultados obtenidos en el desempeño del modelo para predicción de DM2 basado en FCM y una comparación cuantitativa con investigaciones reportadas en la literatura que emplearon el mismo conjunto de datos.

3.1 Caracterización de los individuos

En el análisis descriptivo del conjunto de datos, el 100 % de los individuos eran mujeres (35 % eran diabéticas y el 65 % no diabéticas) con edades comprendidas entre 21 y 81 años, con un promedio de 33 ± 12 años. Las mujeres con diabetes presentaron una edad promedio de 37 ± 11 años y las mujeres sin la enfermedad una media de $31 \pm 11,7$ años. La edad constituye una variable de gran interés en el diagnóstico de la DM2, en [26] encontraron una relación importante entre la edad y el desarrollo de complicaciones, demostrando que al realizar el diagnóstico de la DM2 en edades más tempranas se asoció con un mayor riesgo relativo de morbilidad y mortalidad cardiovascular. En la Tabla 3, se pueden observar los resultados de la media y desviación estándar para cada variable de la clase 0 y clase 1.

Tabla 3. Media y desviación estándar de las variables del conjunto de datos. Fuente: elaboración propia.

Variable	Mujeres sin diabetes (clase 0) n = 500		Mujeres diabéticas (clase 1) n = 268	
	Media	DE	Media	DE
C1	3,3	3,0	4,9	3,0
C2	110,6	24,8	142,3	24,8
C3	70,9	12,2	75,3	12,2
C4	27,2	10,0	33,0	10,0
C5	130,3	102,5	206,8	102,5
C6	30,9	6,6	35,4	6,6
C7	0,4	0,3	0,6	0,3
C8	31,2	11,7	37,1	11,7

3.2 Desempeño predictivo del modelo

El modelo final optimizado consistió en un FCM entrenado con AG usando los siguientes hiperparámetros: población inicial = 200, probabilidad de cruce = 0.5, probabilidad de mutación = 0.5, función de activación = sigmoidea, función de inferencia = rescaled. El modelo obtuvo una exactitud del 99 %, una precisión de 98 % y un recall del 100 % en la predicción de DM2. Con relación al grado de concordancia entre la predicción de nuestro modelo y el diagnóstico real, se obtuvo un índice Kappa de Cohen de 0,98 un valor casi perfecto según [27] que indica que la fiabilidad del modelo es alta. En la Tabla 4, se puede observar la matriz de confusión, que muestra los resultados de la predicción.

Tabla 4. Matriz de confusión obtenida en la evaluación del modelo FCM. Fuente: elaboración propia.

Predicción con FCM-AG		Diagnóstico	
		Sin diabetes	Con diabetes
	Negativa	147	3
	Positiva	0	150

En la Tabla 5, se muestran las métricas de desempeño obtenidas por el modelo y una comparación con los reportes en la literatura. La exactitud indica que el modelo clasificó correctamente el 99 % de las predicciones, la precisión demostró que el 98 % de las pacientes clasificadas como diabéticas son realmente positivas y un recall de 100 % indica que el modelo tiene una buena capacidad para identificar los casos verdaderamente positivos de DM2, lo que demuestra que en general el enfoque propuesto es confiable y presenta un buen desempeño.

Tabla 5: Comparación cuantitativa con modelos reportados en la literatura. Fuente: elaboración propia.

Autores	Modelo	Exactitud	Precisión	Recall
Gupta y Goel [28]	RF	0,88	1,00	0,60
Febrian y colaboradores [29]	NB	0,76	0,73	0,71
Reza y colaboradores [30]	SVM	0,85	0,83	0,87
Tasin y colaboradores [6]	XGBoost	0,81	0,81	0,81
Li y colaboradores [31]	KNN + GA	0,88	-	0,89
Olisah [32]	2GDNN	0,97	0,97	0,97
Nuestro modelo	FCM + GA	0,99	0,98	1,00

En los últimos años, se ha publicado una diversidad de artículos de investigación que emplean técnicas de IA en el conjunto de datos de PIMA para la predicción de la diabetes. Los modelos de ML han sido los más utilizados, presentando una buena capacidad para clasificar y predecir la enfermedad. Una comparativa de las métricas se puede observar en la Tabla 5. Por ejemplo, en [28], realizaron un estudio para comparar el rendimiento de los algoritmos de clasificación en la predicción de la diabetes: K-vecinos más cercanos (KNN), DT, RF y SVM, con diferentes formas de preprocesamiento. El mejor modelo fue RF con una exactitud de 88 % y una precisión de 100 %. RF obtuvo una menor exactitud que el modelo de FCM propuesto, el rendimiento de los clasificadores se ve afectado por los valores incompletos y desconocidos, por lo tanto, para mejorar el desempeño se deben eliminar esos datos. Los autores [28] se enfocan exclusivamente en las técnicas de preprocesamiento y selección de hiperparámetros para desarrollar el mejor modelo predictivo, centrando su trabajo en aspectos matemáticos y computacionales. En comparación, el modelo de FCM desarrollado está orientado al ámbito clínico, integrando el comportamiento y las interacciones de las variables en la predicción de la DM2.

En [29], compararon los algoritmos de KNN y NB para predecir la diabetes. Realizaron 8 experimentos, dividiendo los datos en diferentes porcentajes (del 10 % a 80 %) para el entrenamiento de los modelos. El mejor resultado se obtuvo con el 80 % de los datos, donde el algoritmo NB superó a KNN, con una exactitud de 76,0 %, precisión de 73,3 % y recall de 71,3 %. En [29] se enfocaron en evaluar la predicción de la enfermedad con diferentes porcentajes de datos, sin embargo, no describen la importancia de las variables asociadas a la predicción de la DM2. Los algoritmos de NB pueden presentar limitaciones en el manejo de relaciones complejas, debido a que se basan en la hipótesis de independencia entre las variables predictoras, lo que disminuiría la capacidad para analizar el comportamiento de los factores de riesgo en el desarrollo de la enfermedad. La DM2 puede estar influenciada por múltiples factores, por lo tanto, podría ocasionar la disminución del rendimiento de los modelos de NB. En [30] implementaron un *kernel* no lineal para el algoritmo SVM para mejorar la clasificación de la DM2. El nuevo *kernel* utiliza la función de base radial (RBF) que permiten a SVM aprender límites de decisión complejos y adaptarse al conjunto de datos PIMA. Compararon con varios *kernel* existentes y lograron un rendimiento mejorado con una exactitud de 85,5 %, recall de 87,0 % y precisión de 83,4 %. En [6], realizaron un estudio

utilizando varios modelos de aprendizaje automático, aplicaron técnicas de balance de datos como SMOTE y *Adaptive Synthetic Sampling Approach for Imbalanced Learning* (ADASYN, por sus siglas en inglés). Los mejores resultados fueron obtenidos por el clasificador XGBoost con una exactitud de 81 % con el enfoque ADASYN.

En [31], utilizaron diferentes algoritmos metaheurísticos para mejorar el diagnóstico de la diabetes. Examinaron varias combinaciones del algoritmo de búsqueda Harmony, AG y PSO con K-means para la selección de características. El modelo de AG con KNN obtuvo una exactitud de 88,02 %. En [32], propusieron un *framework* de aprendizaje automático para mejorar el rendimiento de la predicción de la diabetes utilizando los conjuntos de datos PIMA y Laboratory of the Medical City Hospital (LMCH, por sus siglas en inglés). El *framework* incorporó enfoques de preprocesamiento de datos, correlación de Spearman y regresión polinomial. El mejor modelo propuesto por los autores fue una red neuronal profunda de doble crecimiento (2GDNN) con una exactitud, precisión y recall de 97 %. Este enfoque presenta algunas limitaciones, por ejemplo, que requieren grandes cantidades de datos para su entrenamiento y son difíciles de interpretar, lo que complica entender la forma en cómo se realiza la predicción [33].

Al comparar los resultados del modelo propuesto en esta investigación con otros enfoques reportados en la literatura, se puede evidenciar que el FCM optimizado con AG presenta un excelente desempeño para predecir la DM2. El modelo de FCM que se propone en esta investigación presentó un rendimiento superior en comparación con los trabajos de [6] y [30], ambos con un desempeño alrededor del 80 %. Probablemente se deba a que los clasificadores usados en los estudios previos requieren grandes conjuntos de datos para su entrenamiento, lo que resulta limitado con la base de datos de PIMA. Además, estos algoritmos suelen ser difíciles de interpretar y requieren el uso de otras herramientas para ser explicables, lo que disminuiría su aplicación como sistemas de apoyo a la decisión médica. Las ventajas de la implementación de FCM con AG están respaldadas por la facilidad que tienen los FCM para representar la influencia de las variables relacionadas con la diabetes. Los FCM tienen en cuenta las influencias positivas y negativas, lo que permite mejorar la comprensión del análisis de los factores de riesgos asociados a la enfermedad. Otra ventaja, es la capacidad de realizar procesos iterativos, lo que apoya en el análisis del comportamiento de las variables a través del tiempo, por lo tanto, son una buena estrategia para la toma de decisiones médicas. Además, el uso de representaciones gráficas permite a los expertos visualizar y razonar sobre las predicciones del modelo, facilitando la comprensión del sistema desarrollado [19].

3.3 Simulación de las variables usadas en el modelo FCM

Una aplicación importante de los FCM es poder ser utilizados como modelos de apoyo a la toma de decisiones clínicas, por su facilidad en la interpretación. Permiten la generación de gráficas visualizables para el análisis de las variables asociadas al desarrollo de la enfermedad, permitiendo predecir el comportamiento de los conceptos en el tiempo [10].

Analizar el comportamiento de los factores de riesgo, es una de las prioridades para la detección temprana de la diabetes. El modelo de FCM optimizado con AG identificó relaciones positivas entre el diagnóstico de DM2 y los conceptos: C1 (número de embarazos), C2 (glucosa plasmática), C3 (presión sanguínea diastólica), C5 (insulina sérica) y C7 (función de pedigrí). Estudios han reportado que las mujeres con antecedentes de múltiples embarazos pueden tener un mayor riesgo de desarrollar la enfermedad, principalmente, si experimentaron diabetes gestacional [34], [35]. La concentración plasmática de glucosa e insulina están directamente relacionadas con el diagnóstico de la DM2 [36]. Las personas con niveles

elevados de glucosa tienen una mayor probabilidad de adquirir la patología. Estos valores están relacionados con las deficiencias en la secreción de insulina por las células β pancreáticas y con la resistencia a la hormona por los tejidos periféricos, por lo tanto, son factores clave en el desarrollo e identificación de la diabetes [37]. Otro concepto asociado, es la presencia de hipertensión arterial, que ocurre en el 50 % al 80 % de los pacientes con DM2. Estas variables tienen una relación de causalidad recíproca, debido a que la resistencia a la insulina y la diabetes predisponen a la hipertensión, la rigidez vascular y la enfermedad cardiovascular. Mientras que, la rigidez arterial y la vasorelajación alterada contribuyen a empeorar la resistencia a la insulina y al desarrollo de la diabetes [38]. Con relación a la función pedigrí, proporciona información relevante al utilizar los antecedentes clínicos familiares para predecir la probabilidad de desarrollar diabetes [39]. Algunos autores reportaron otras variables de interés en el mismo conjunto de datos; por ejemplo, en [40] identificaron que los cinco conceptos críticos para predecir la DM2 son la edad, el índice de masa corporal (IMC), la función pedigrí, la glucosa y la frecuencia de embarazos, a través de un modelo de regresión logística. En [41] reportaron que el IMC y la glucosa plasmática son atributos significativamente correlacionados con la predicción de la enfermedad.

En la Figura 4, se observa una representación gráfica del comportamiento de las variables predictoras de DM2. La simulación utiliza a una paciente con antecedentes de embarazo (línea amarilla), presión diastólica alterada (línea verde) e IMC aumentado (línea violeta). Las líneas en el gráfico representan la evolución de los conceptos a lo largo de las 35 iteraciones, donde los valores cercanos a 1 indican una mayor importancia de las variables asociadas. Algunos conceptos que inicialmente se encontraban inactivos, muestran un incremento a medida que avanza la simulación, como la concentración de glucosa (línea roja), niveles de insulina sérica (línea naranjada) y espesor del pliegue cutáneo del tríceps (línea morada). Al inicio no estaban activos, pero, posteriormente aparecen como curvas que incrementan sus valores a lo largo del tiempo.

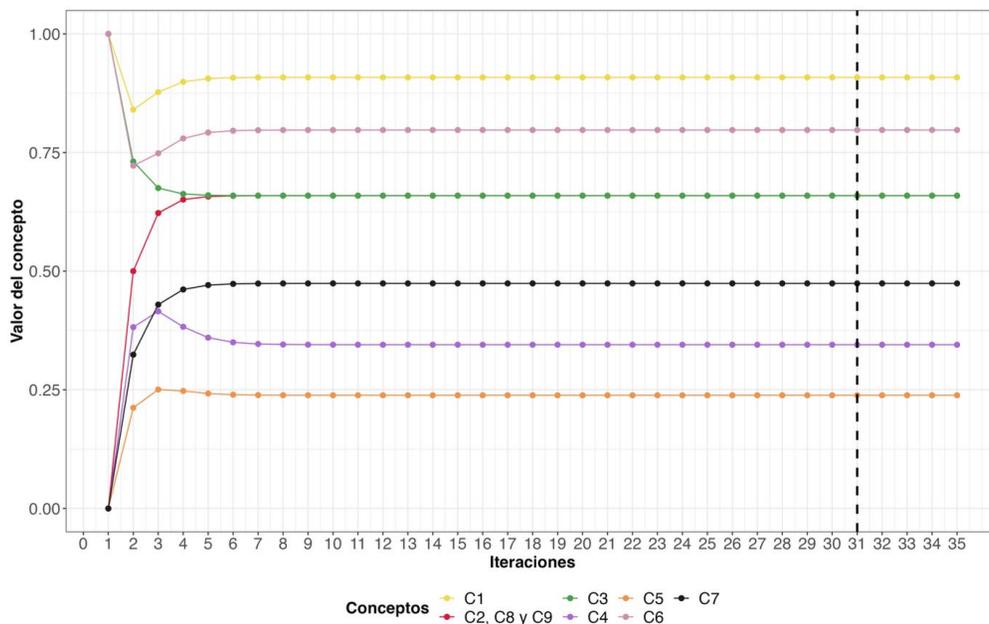


Figura 4. Representación gráfica del comportamiento de las variables predictoras de DM2.

Fuente: elaboración propia.

De esta manera, estos resultados representados en forma gráfica le permiten al profesional médico realizar un análisis más completo de la DM2 para cada paciente. Los FCM

pueden analizar variables sociodemográficas (edad, sexo, etc.), variables clínicas (signos, síntomas, historial médico, etc.) y resultados de mediciones biológicas sobre los pacientes (pruebas de laboratorio, índices biológicos, entre otros). Al visualizar estas relaciones, los FCM pueden ayudar a los médicos a predecir la progresión de la enfermedad, identificar el comportamiento de los factores de riesgo y realizar intervenciones tempranas que permitan tomar decisiones informadas al proporcionar una comprensión más completa de la condición del paciente [42].

4. CONCLUSIONES

La DM2 es una enfermedad metabólica crónica que puede afectar diversos órganos y está influenciada por los estilos de vida. Las altas tasas de morbilidad y mortalidad han centrado la atención de la aplicación de técnicas de IA. En este estudio, se propuso un modelo de predicción inteligente de DM2 basado en FCM optimizado con AG, con el objetivo de analizar el comportamiento de las variables que tienen un riesgo sustancial para predecir la diabetes en mujeres. Los resultados demuestran el potencial de los FCM para la predicción de la DM2 con un excelente desempeño, al lograr una exactitud del 99 %, precisión de 98 % y recall de 100 %, valores superiores a los reportados en la literatura. El modelo permite el análisis del comportamiento de las variables predictoras, al generar gráficos visualizables e interpretables de las variables en el tiempo. Este aspecto lo convierte en una opción para ser utilizado como sistema de apoyo a la toma de decisiones médicas, especialmente para profesionales con experiencia limitada y en zonas con acceso escaso a los sistemas de salud especializados. De esta manera se podría contribuir en la toma de acciones que prevengan el inicio de la enfermedad.

El presente estudio presenta algunas limitaciones. Inicialmente, están relacionadas con el alcance de los datos, debido a que la muestra sólo contiene información de mujeres de la etnia Pima, esto podría limitar la capacidad del modelo para generalizarse en otras poblaciones. Los factores de riesgo pueden variar entre las zonas geográficas, lo que podría disminuir la eficacia del modelo si se aplica en otra población. Por lo tanto, aumentar el número de registros, al incluir datos masculinos o de otros grupos sociales permitiría mejorar el análisis de la enfermedad. Para trabajos futuros, sería de interés emplear otras variables como pruebas de laboratorio, factores sociodemográficos, económicos y estilos de vida, importantes para comprender el riesgo de desarrollar DM2. Además, sería pertinente involucrar expertos en el área que asignen relaciones entre los conceptos de FCM. También incluir datos longitudinales para verificar el comportamiento de las variables a lo largo del tiempo. Finalmente, otro trabajo futuro interesante sería comparar el uso de los FCM con otros modelos de ML.

5. AGRADECIMIENTO Y FINANCIACIÓN

Los autores agradecen a los creadores del conjunto de datos. Adicionalmente, agradecen a la Universidad Cooperativa de Colombia por el apoyo parcial a esta investigación bajo el proyecto INV3569.

CONFLICTOS DE INTERÉS DE LOS AUTORES

Los autores declaran que no tienen conflictos de interés en competencia, ni relaciones personales conocidas que pudieran haber influido en el desarrollo del trabajo presentado en este artículo.

CONTRIBUCIÓN DE LOS AUTORES

William Hoyos: conceptualización, metodología, software, análisis, investigación, curación de datos, validación, visualización y escritura final de artículo.

Rander Ruiz: conceptualización, curación de datos, supervisión, redacción, revisión y edición.

Kenia Hoyos: conceptualización, metodología, análisis, validación, supervisión, escritura, revisión y edición.

6. REFERENCIAS

- [1] American Diabetes Association Professional Practice Committee *et al.*, “2. Diagnosis and classification of diabetes: *standards of care in diabetes—2024*,” *Diabetes Care*, vol. 47, no. Supplement_1, pp. S20–S42, 2024. <https://doi.org/10.2337/dc24-S002>
- [2] World Health Organization. “Diabetes.” who.int. Accessed: Feb. 1, 2024. [Online.] Available: <https://www.who.int/news-room/fact-sheets/detail/diabetes>
- [3] D. J. Magliano, E. J. Boyko, and IDF Diabetes Atlas 10th edition scientific committee, *IDF DIABETES ATLAS*, Brussels: International Diabetes Federation, 2021. <http://www.ncbi.nlm.nih.gov/books/NBK581934/>
- [4] C. Bommer *et al.*, “The global economic burden of diabetes in adults aged 20–79 years: a cost-of-illness study,” *Lancet Diabetes Endocrinol.*, vol. 5, no. 6, pp. 423-430, Jun. 2017. [https://doi.org/10.1016/S2213-8587\(17\)30097-9](https://doi.org/10.1016/S2213-8587(17)30097-9)
- [5] J. Zhang *et al.*, “The impact of global, regional, and national population ageing on disability-adjusted life years and deaths associated with diabetes during 1990–2019: A global decomposition analysis,” *Diabetes Metab. Syndr. Clin. Res. Rev.*, vol. 17, no 6, p. 102791, Jun. 2023. <https://doi.org/10.1016/j.dsx.2023.102791>
- [6] I. Tasin, T. U. Nabil, S. Islam, and R. Khan, “Diabetes prediction using machine learning and explainable AI techniques,” *Healthc. Technol. Lett.*, vol. 10, no. 1–2, pp. 1–10, Dec. 2023. <https://doi.org/10.1049/htl2.12039>
- [7] J. Abdollahi, and B. Nouri-Moghaddam, “Hybrid stacked ensemble combined with genetic algorithms for diabetes prediction,” *Iran J. Comput. Sci.*, vol. 5, no 3, pp. 205-220, Sep. 2022. <https://doi.org/10.1007/s42044-022-00100-1>
- [8] A. Hennebelle, H. Materwala, and L. Ismail, “HealthEdge: A Machine Learning-Based Smart Healthcare Framework for Prediction of Type 2 Diabetes in an Integrated IoT, Edge, and Cloud Computing System,” *Procedia Comput. Sci.*, vol. 220, pp. 331-338, 2023. <https://doi.org/10.1016/j.procs.2023.03.043>
- [9] T. Mora, D. Roche, and B. Rodríguez-Sánchez, “Predicting the onset of diabetes-related complications after a diabetes diagnosis with machine learning algorithms,” *Diabetes Res. Clin. Pract.*, vol. 204, p. 110910, Oct. 2023. <https://doi.org/10.1016/j.diabres.2023.110910>
- [10] A. Amirkhani, E. I. Papageorgiou, A. Mohseni, and M. R. Mosavi, “A review of fuzzy cognitive maps in medicine: Taxonomy, methods, and applications,” *Comput. Methods Programs Biomed.*, vol. 142, pp. 129-145, Apr. 2017. <https://doi.org/10.1016/j.cmpb.2017.02.021>
- [11] A. M. Shukla, P. Pandit, V. Purandare, and A. Srinivasaraghavan, “Symptom based Hierarchical Classification of Diabetes and Thyroid disorders using Fuzzy Cognitive Maps,” *Cornell University*, no. 2, 2021. <https://doi.org/10.48550/arXiv.2108.03760>
- [12] A. Alam, “Fuzzy Cognitive Maps Approach to Identify Risk Factors of Diabetes,” *Journal of Physical Sciences*, vol. 22, pp. 13-21, 2017. <https://vixra.org/pdf/1803.0471v1.pdf>
- [13] W. Hoyos, K. Hoyos, and R. Ruiz-Pérez, “Modelo de inteligencia artificial para la detección temprana de diabetes,” *Biomédica*, vol. 43, no. Sp.3, pp. 110-121, Dec. 2023. <https://doi.org/10.7705/biomedica.7147>
- [14] S. Almufti, A. Yahya, and H. Khalid, “A comparative study of particle swarm optimization and genetic

- algorithm,” *International Journal of Electrical Engineering*, vol. 9, no. 2, pp. 215–223, 2016. http://www.irphouse.com/ijee16/ijeev9n2_06.pdf
- [15] J. W. Smith, J. E. Everhart, W. C. Dickson, W. C. Knowler, and R. S. Johannes, “Using the ADAP learning algorithm to forecast the onset of diabetes mellitus,” *Proceedings of the Annual Symposium on Computer Application in Medical Care*, p. 261, Nov.1988. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2245318/>
- [16] UCI Machine Learning, and Kaggle Team “Pima Indians Diabetes Database.” Kaggle.com. Accessed: Feb. 1, 2024. [Online.] Available: <https://www.kaggle.com/datasets/uciml/pima-indians-diabetes-database>
- [17] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, “SMOTE: Synthetic Minority Over-sampling Technique,” *J. Artif. Intell. Res.*, vol. 16, pp. 321-357, Jun. 2002. <https://doi.org/10.1613/jair.953>
- [18] B. Kosko, “Fuzzy cognitive maps,” *Int. J. Man-Mach. Stud.*, vol. 24. no. 1, pp. 65-75, Jan. 1986. [https://doi.org/10.1016/S0020-7373\(86\)80040-2](https://doi.org/10.1016/S0020-7373(86)80040-2)
- [19] I. D. Apostolopoulos, N. I. Papandrianos, N. D. Papathanasiou, and E. I. Papageorgiou, “Fuzzy Cognitive Map Applications in Medicine over the Last Two Decades: A Review Study,” *Bioengineering*, vol. 11, no. 2, p. 139, Jan. 2024. <https://doi.org/10.3390/bioengineering11020139>
- [20] W. Hoyos, J. Aguilar, and M. Toro, “A clinical decision-support system for dengue based on fuzzy cognitive maps,” *Health Care Manag. Sci.*, vol. 25, no. 4, pp. 666-681, Dec. 2022. <https://doi.org/10.1007/s10729-022-09611-6>
- [21] B. Alhijawi, and A. Awajan, “Genetic algorithms: theory, genetic operators, solutions, and applications,” *Evol. Intell.*, vol. 17, no. 3 pp. 1245-1256, Feb. 2023. <https://doi.org/10.1007/s12065-023-00822-6>
- [22] S. Katoch, S. S. Chauhan, and V. Kumar, “A review on genetic algorithm: past, present, and future,” *Multimed. Tools Appl.*, vol. 80, no. 5, pp. 8091-8126, Feb. 2021. <https://doi.org/10.1007/s11042-020-10139-6>
- [23] A. Choudhary, M. Kumar, M. K. Gupta, D. K. Unune, and M. Mia, “Mathematical modeling and intelligent optimization of submerged arc welding process parameters using hybrid PSO-GA evolutionary algorithms,” *Neural Comput. Appl.*, vol. 32, no. 10, pp. 5761–5774, May. 2020. <https://doi.org/10.1007/s00521-019-04404-5>
- [24] D. Liu, “Mathematical modeling analysis of genetic algorithms under schema theorem,” *J. Comput. Methods Sci. Eng.*, vol. 19, pp. 131-137, Aug. 2019. <https://doi.org/10.3233/JCM-191019>
- [25] H. Wickham, "Toolbox," in *ggplot2 Elegant Graphics for Data Analysis*, Cham: Springer International Publishing, 2016, ch. 3, Toolbox, pp. 33-74. <https://doi.org/10.1007/978-3-319-24277-4>
- [26] D. H. Seo *et al.*, “Association between age at diagnosis of type 2 diabetes and cardiovascular morbidity and mortality risks: A nationwide population-based study,” *Diabetes Res. Clin. Pract.*, vol. 208, p. 111098, Feb. 2024. <https://doi.org/10.1016/j.diabres.2024.111098>
- [27] J. R. Landis, and G. G. Koch, “The measurement of observer agreement for categorical data,” *Biometrics*, vol. 33, no. 1, pp. 159-174, Mar. 1977. <https://doi.org/10.2307/2529310>
- [28] S. C. Gupta, and N. Goel, “Predictive Modeling and Analytics for Diabetes using Hyperparameter tuned Machine Learning Techniques,” *Procedia Comput. Sci.*, vol. 218, pp. 1257-1269, 2023. <https://doi.org/10.1016/j.procs.2023.01.104>
- [29] M. E. Febrian, F. X. Ferdinan, G. P. Sendani, K. M. Suryanigrum, and R. Yunanda, “Diabetes prediction using supervised machine learning,” *Procedia Comput. Sci.*, vol. 216, pp. 21-30, 2023. <https://doi.org/10.1016/j.procs.2022.12.107>
- [30] Md. S. Reza, U. Hafsha, R. Amin, R. Yasmin, and S. Ruhi, “Improving SVM performance for type II diabetes prediction with an improved non-linear kernel: Insights from the PIMA dataset,” *Comput. Methods Programs Biomed. Update*, vol. 4, p. 100118, 2023. <https://doi.org/10.1016/j.cmpbup.2023.100118>
- [31] X. Li, J. Zhang, and F. Safara, “Improving the Accuracy of Diabetes Diagnosis Applications through a Hybrid Feature Selection Algorithm,” *Neural Process. Lett.*, vol. 55, no. 1, pp. 153-169, Feb. 2023. <https://doi.org/10.1007/s11063-021-10491-0>
- [32] C. C. Olisah, L. Smith, and M. Smith, “Diabetes mellitus prediction and diagnosis from a data preprocessing and machine learning perspective,” *Comput. Methods Programs Biomed.*, vol. 220, p. 106773, Jun. 2022. <https://doi.org/10.1016/j.cmpb.2022.106773>
- [33] G. Dharmarathne, T. N. Jayasinghe, M. Bogahawaththa, D. P. P. Meddage, and U. Rathnayake, “A novel machine learning approach for diagnosing diabetes with a self-explainable interface,” *Healthc. Anal.*, vol. 5, p. 100301, Jun. 2024. <https://doi.org/10.1016/j.health.2024.100301>
- [34] A-M. Auvinen *et al.* “Type 1 and type 2 diabetes after gestational diabetes: a 23 year cohort study,” *Diabetologia*, vol 63, no. 10, pp. 2123–2128, Oct. 2020. <https://doi.org/10.1007/s00125-020-05215-3>
- [35] E. Hanson, I. Ringmets, A. Kirss, M. Laan, and K. Rull, “Screening of gestational diabetes and its risk factors: Pregnancy outcome of women with gestational diabetes risk factors according to glycose tolerance test results,” *J. Clin. Med.*, vol. 11, no. 17, p. 4953, Aug. 2022. <https://doi.org/10.3390/jcm11174953>
- [36] Organización Panamericana de la Salud, “Diagnóstico y manejo de la diabetes de tipo 2,” *Ops*, pp. 1–38, 2020. Disponible en:

- https://iris.paho.org/bitstream/handle/10665.2/53007/OPSWNMHNV200043_spa.pdf?sequence=1&isAllowed=y
- [37] N. Rachdaoui, “Insulin: The friend and the foe in the development of type 2 diabetes mellitus,” *Int. J. Mol. Sci.*, vol. 21, no. 5, p. 1770, Mar. 2020. <https://doi.org/10.3390/IJMS21051770>
- [38] G. Jia, and J. Sowers, “Hypertension in Diabetes: An Update of Basic Mechanisms and Clinical Disease,” *Hypertension*, vol. 78, no. 5, pp. 1197–1205, Oct. 2021. <https://doi.org/10.1161/HYPERTENSIONAHA.121.17981>
- [39] M. Das *et al.*, “Determinants of gestational diabetes pedigree function for Pima Indian females,” *Intern. Med. - Open J.*, vol. 6, no. 1, pp. 9–13, Dec. 2022. <https://doi.org/10.17140/imoj-6-121>
- [40] R. D. Joshi, and C. K. Dhakal, “Predicting type 2 diabetes using logistic regression and machine learning approaches,” *Int. J. Environ. Res. Public Health*, vol. 18, no. 14, p. 7346, Jul. 2021. <https://doi.org/10.3390/ijerph18147346>
- [41] R. Birjais, A. K. Mourya, R. Chauhan, and H. Kaur, “Prediction and diagnosis of future diabetes risk: a machine learning approach,” *SN Appl. Sci.*, vol. 1, no. 9, Aug. 2019. <https://doi.org/10.1007/s42452-019-1117-9>
- [42] I. D. Apostolopoulos, P. P. Groumpos, and D. J. Apostolopoulos, “Advanced fuzzy cognitive maps: state-space and rule-based methodology for coronary artery disease detection,” *Biomed. Phys. Eng. Express*, vol. 7, no. 4, p. 045007, 2021. <https://doi.org/10.1088/2057-1976/abfd83>