

# Predicción de diabetes usando redes neuronales multicapa

Brandon Nahú Nava-Martínez, Juan Carlos Elizondo-Leal

*Facultad de Ingeniería y Ciencias, Universidad Autónoma de Tamaulipas, Ciudad Victoria 87000, México*

**Resumen**—En este estudio, se ha implementado y evaluado un clasificador de red neuronal multicapa (MLP) para predecir la presencia de diabetes utilizando un conjunto de datos estándar. El modelo alcanzó una precisión del 84.42% y superó en precisión a varios métodos tradicionales de clasificación. Se presentan los detalles del preprocesamiento de los datos, la estructura del modelo y los resultados obtenidos.

## I. INTRODUCCIÓN

La salud pública es crucial para proteger a la comunidad y prevenir enfermedades peligrosas [1]. La diabetes mellitus es una de las enfermedades extremadamente mortales porque contribuye a otras enfermedades letales, como el daño cardíaco, renal y nervioso [2]. Se estima que afecta a millones de personas y su prevalencia sigue en aumento, lo que representa un desafío significativo para los sistemas de salud pública. La capacidad de predecir y detectar tempranamente la diabetes puede jugar un papel crucial en la implementación de estrategias preventivas y en la reducción de las complicaciones asociadas a la enfermedad.

En el ámbito del aprendizaje automático, las redes neuronales multi-capas (MLP) han demostrado ser herramientas poderosas para la clasificación y predicción en una amplia variedad de dominios. Las MLP son capaces de modelar relaciones complejas y no lineales en los datos, lo que se convierte en una elección adecuada para la tarea de predicción de enfermedades como la diabetes.

Este estudio se centra en la implementación y evaluación de un clasificador MLP para la predicción de diabetes utilizando el conjunto de datos de diabetes de Pima Indian. Este conjunto de datos incluye varias características relevantes que pueden influir en la aparición de la diabetes, tales como el número de embarazos, los niveles de glucosa, la presión arterial, el grosor de la piel, los niveles de insulina, el índice de masa corporal (IMC), la función de pedigrí de la diabetes y la edad.

El objetivo principal de este estudio es desarrollar un modelo predictivo que pueda distinguir entre individuos diabéticos y no diabéticos con alta precisión. Para lograr esto, se aplicaron técnicas de preprocesamiento de datos, tales como la codificación de etiquetas y la estandarización de características, para preparar los datos para el modelo MLP. Además, se utilizaron métodos de validación cruzada y ajuste

de hiperparámetros para optimizar el rendimiento del modelo.

En las siguientes secciones, se describen en detalle la metodología utilizada para el preprocesamiento de los datos, la configuración del modelo MLP, y los resultados obtenidos. Finalmente, se discuten las implicaciones de los hallazgos y las posibles direcciones futuras para mejorar la precisión del modelo y su aplicabilidad en la práctica clínica.

## II. METODOLOGÍA

### A. Conjunto de Datos

El conjunto de datos utilizado en este estudio se obtiene del conjunto de datos de diabetes de Pima Indian. Contiene 768 instancias y 9 características: 'Pregnancies', 'Glucose', 'BloodPressure', 'SkinThickness', 'Insulin', 'BMI', 'DiabetesPedigreeFunction', 'Age' y 'Outcome'.

### B. Preprocesamiento de Datos

Se realizaron las siguientes operaciones de preprocesamiento:

1. Codificación de etiquetas: Se aplicó LabelEncoder a la columna de resultados ('Outcome').
2. Estandarización de características: Se utilizaron StandardScaler para escalar las características.

### C. Modelo de Clasificación

Existen diferentes tipos de modelos de machine learning de los cuales se eligió el modelo Multi-Layer Perceptron (MLP), el cual es un tipo de red neuronal artificial de alimentación hacia adelante que se compone de al menos tres capas: una capa de entrada, una o más capas ocultas y una capa de salida.

Este tipo de modelo tiene los siguientes componentes clave:

- Capas de Entrada: Recibe los datos de entrada. El número de neuronas en esta capa es igual al número de características en los datos de entrada.
- Capas Ocultas: Procesan las señales de entrada a través de múltiples capas. Cada capa oculta contiene un conjunto de neuronas que realizan operaciones de ponderación y activación.
- Capa de Salida: Genera el resultado final de la red, en función de la aplicación (clasificación, regresión, etc.). el número de neuronas en esta capa es igual al número de clases o valores de salida.

\* Revista Argentina de Trabajos Estudiantiles. Patrocinada por la IEEE.

### Funcionamiento del MLP:

- Ponderación: Cada conexión entre neuronas tiene un peso asociado que se ajusta durante el entrenamiento para minimizar el error de predicción.
- Función de Activación: Después de ponderar las entradas, se aplica una función de activación (como ReLU, Sigmoid, Tanh) para introducir no linealidad en el modelo.
- Entrenamiento: Utiliza algoritmos de optimización como el gradiente descendente para ajustar los pesos, minimizando una función de pérdida basada en la diferencia entre la predicción y el valor real.
- Backpropagation: Un método para ajustar los pesos de la red minimizando el error de predicción. Propaga el error desde la capa de salida hacia atrás a través de las capas ocultas hasta la capa de entrada.

En este artículo se utilizó una red neuronal multi-capas (MLP) con la siguiente configuración:

- Capas ocultas: (11, 11)
- Máximo de iteraciones: 700
- Función de activación: ReLU
- Optimizador: SGD
- Tasa de aprendizaje: 0.001
- Alpha (regularización): 0.0001

### III. RESULTADOS

Para generar el modelo de clasificación se utilizaron los hiperparámetros de la Tabla I. El rendimiento del modelo se evaluó utilizando las métricas Accuracy, Precision, Recall, F1-score. Estas métricas se definen como:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

$$F1\_score = \frac{2 \times Precision \times Recall}{Recall + Precision} \quad (4)$$

Donde TP denota el número de verdaderos positivos, TN denota el número de verdaderos negativos, FN denota el número de falsos negativos, FP denota el número de falsos positivos.

TABLA I  
HIPERPARÁMETROS

Parámetro	Valor
Capas ocultas	(11, 11)
Máximo de iteraciones	700
Función de activación	ReLU
Optimizador	SGD
Tasa de aprendizaje	0.001
Alpha	0.0001

Los resultados indican que el modelo MLP alcanzó una precisión del 84.42%. La precisión y el recall del modelo sugieren un desempeño sólido, aunque con diferencias entre clases.

TABLA II  
RESULTADOS DE LA CLASIFICACIÓN

	Precision	ReCall	F1-score	Support
0	0.860	0.911	0.885	101
1	0.809	0.717	0.760	53
Accuracy			0.844	154
Macro avg	0.834	0.814	0.822	154
Weighted avg	0.842	0.844	0.842	154

El modelo mostró un rendimiento balanceado en términos de precisión y recall, siendo particularmente efectivo en la clasificación de la clase negativa (no diabetes). Sin embargo, el recall de la clase positiva (diabetes) es más bajo, indicando algunos falsos negativos.

TABLA II  
MATRIZ DE CONFUSION DEL MODELO MLP

### IV. DISCUSIÓN

El modelo MLP superó en rendimiento a varios enfoques

	Negativo	Positivo
Falso	92	9
Verdadero	15	38

tradicionales de clasificación en términos de exactitud y F1-score. No obstante, es necesario trabajar en la mejora de la sensibilidad del modelo hacia la clase positiva para reducir la tasa de falsos negativos.

Estas métricas y resultados resaltan el potencial del MLP para la predicción de diabetes, con espacio para ajustes y mejoras adicionales.

### V. CONCLUSIÓN

El uso de redes neuronales multi-capas para la predicción de diabetes ha demostrado ser una técnica efectiva. Sin embargo, se debe considerar la mejora de la precisión para la clase de diabetes mediante el ajuste de hiperparámetros y el uso de técnicas adicionales de preprocesamiento de datos.

### REFERENCIAS

- [1] World Health Organization, Global Action Plan on Physical Activity 2018-2030: More Active People for a Healthier World, World Health Organization, Geneva, Switzerland, 2019. J. Clerk Maxwell, *A Treatise on Electricity and Magnetism*, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68-73.
- [2] American Diabetes Association, "Diagnosis and classification of diabetes mellitus," *Diabetes Care*, vol. 37, no. Supplement 1, pp. S81-S90, 2014.