



UNIVERSIDAD TECNOLÓGICA CENTROAMERICANA

FACULTAD DE INGENIERÍA

PROYECTO DE INVESTIGACIÓN

**DESARROLLO DE RED NEURONAL ARTIFICIAL PARA EL DIAGNÓSTICO CLÍNICO Y PROBABILÍSTICO
DE DIABETES MELLITUS**

**PREVIO A LA OBTENCIÓN DEL TÍTULO
INGENIERO EN BIOMÉDICA**

PRESENTADO POR:

21911117 JENNIFER JEANNETTE GARCÍA GUILLÉN

21911249 MIRIAM ZULEMA BADAY TORREZ

ASESOR METODOLÓGICO: MANUEL GAMERO

ASESOR TEMÁTICO: DELOND JIMÉNEZ

SAN PEDRO SULA, CORTÉS, HONDURAS, C.A.

ABRIL 2023

AUTORIZACIÓN

AUTORIZACIÓN DE LOS AUTORES PARA LA CONSULTA, LA REPRODUCCIÓN PARCIAL O TOTAL, Y PUBLICACIÓN ELECTRÓNICA DEL TEXTO COMPLETO DE TESIS DE GRADO.

Señores

CENTRO DE RECURSOS PARA EL APRENDIZAJE Y LA INVESTIGACION (CRAI)
San Pedro Sula

Estimados Señores:

La presentación del documento de tesis forma parte de los requerimientos y procesos establecidos de graduación para alumnos de pregrado de UNITEC.

Nosotros, Miriam Zulema Baday Torrez y Jennifer Jeannette García Guillén de San Pedro Sula, autores del trabajo de grado titulado: Desarrollo de una red neuronal artificial para el diagnóstico clínico y probabilístico de diabetes mellitus, presentado y aprobado en el año 2023, como requisito para optar al título de Profesional de Ingeniero Biomédico, autorizo a:

Las Bibliotecas de los Centros de Recursos para el Aprendizaje y la Investigación (CRAI) de la Universidad Tecnológica Centroamericana (UNITEC), para que, con fines académicos, pueda libremente registrar, copiar y usar la información contenida en él, con fines educativos, investigativos o sociales de la siguiente manera:


Los usuarios puedan consultar el contenido de este trabajo de grado en las salas de estudio de la biblioteca y la página Web de la universidad.

Permita la consulta y la reproducción, a los usuarios interesados en el contenido de este trabajo, para todos los usos que tengan finalidad académica, ya sea en formato CD o digital desde Internet, Intranet, etc., y en general para cualquier formato conocido o por conocer.

De conformidad con lo establecido en el artículo 19 de la Ley de Derechos de Autor y de los Derechos Conexos; los cuales son irrenunciables, imprescriptibles, inembargables e inalienables.

Es entendido que cualquier copia o reproducción del presente documento con fines de lucro no está permitida sin previa autorización por escrito de parte de los principales autores.

En fe de lo cual, se suscribe la presente acta en la ciudad de San Pedro Sula a los 28 días del mes de abril de dos mil veintitrés.



Miriam Zulema Baday Torrez
21911249



Jennifer Jeannette García Guillén
21911117

HOJA DE FIRMAS

Los abajo firmantes damos fe, en nuestra posición de miembro de Terna, Asesor y/o Jefe Académico y en el marco de nuestras responsabilidades adquiridas, que el presente documento cumple con los lineamientos exigidos por la Facultad de Ingeniería y los requerimientos académicos que la Universidad dispone dentro de los procesos de graduación.



Ing. Manuel Gamero
Asesor UNITEC



Ing. Juan Carlos Dubón
Miembro 1 de Terna



Ing. William Taylor
Miembro 2 de Terna



The stamp is circular with the text 'UNIVERSIDAD TECNOLÓGICA CENTROAMERICANA' around the top edge. In the center, it says 'unitec' with a logo. Below that, it reads 'JEFATURA ACADÉMICA' and 'INGENIERÍA EN BIOMÉDICA'.

Ing. Reyna Valle

Jefe Académico de Ingeniería en Biomédica | UNITEC

DEDICATORIA

A Dios, a nuestras familias y a nosotras mismas.

AGRADECIMIENTOS

La gratitud es un estilo de vida y las palabras no serán suficientes para expresar lo que sentimos. En esta etapa de nuestras vidas dónde todo termina y todo inicia, agradecemos primeramente a Dios por guiar cada paso de nuestra vida y porque su amor y misericordia nunca nos abandonan.

Agradecemos a nuestros padres, que sin su ayuda, esfuerzo, amor y fe no hubiésemos podido ser lo que ahora somos. Por siempre alentar nuestras ideas y demostrarnos que todo es posible incluso cuando no parece posible. Por ser nuestra fuerza y nuestro motor para avanzar.

A nuestros hermanos y abuelos, por formar parte importante con su apoyo emocional e incondicional, por creer en nosotras y estar presentes siempre.

A nuestros amigos y cercanos que, con su compañía, palabras de aliento y motivación formaron parte indirecta de nuestra actitud durante todo el camino.

A nuestros formadores durante toda nuestra trayectoria académica y a cada una de las personas que marcaron un antes y un después en nuestro crecimiento educativo y personal.

Al ing. Manuel Gamero por su apoyo como asesor metodológico, por creer en nuestro trabajo y motivarnos a ser mejores en el proceso de investigación y a tomar retos desafiantes.

Al ing. Delond Jiménez por su apoyo como asesor temático, por creer en nosotras y por dedicar su tiempo a apoyar nuestro proyecto, por dirigirnos e impulsarnos a mejorar en la programación por medio de Python.

A los establecimientos de salud que nos abrieron las puertas para la recolección de datos y que creyeron en el proyecto, muchas gracias por decir "sí" sin que existiese certeza de un resultado positivo.

- **Jennifer García y Miriam Baday**

EPÍGRAFE

"No necesitamos magia para cambiar el mundo, llevamos todo el poder que necesitamos dentro de nosotros mismos: tenemos el poder de imaginar algo mejor."

-JK Rowling

RESUMEN EJECUTIVO

La Diabetes Mellitus actualmente representa una de las enfermedades crónicas y sistémicas más frecuentes, que genera gran preocupación en el sistema sanitario debido al aumento de casos en los últimos años. Además, la tendencia ha señalado un continuo aumento durante los años siguientes. Ha requerido un diagnóstico precoz para iniciar cambios en el estilo de vida de las personas que padecen la enfermedad con el fin de evitar que la patología evolucione y cause complicaciones al paciente. La inteligencia artificial y el aprendizaje automático han permitido a los médicos disponer de un método de ayuda al diagnóstico para agilizar los procesos y, al mismo tiempo, reducir la cantidad de recursos empleados en el transcurso. Con el objetivo de desarrollar un método de aprendizaje automático que proporcione un diagnóstico clínico y probabilístico, se ha diseñado un modelo de red neuronal artificial en lenguaje de programación Python para analizar datos preclínicos y pruebas de laboratorio recomendadas por expertos. Tras haber evaluado otros 2 métodos de aprendizaje automático, se ha utilizado el árbol de decisión como método de validación debido a su rendimiento y precisión. El diagnóstico brindado por la red neuronal se ha corroborado con el criterio del médico experto y esta ha demostrado tener un rendimiento bastante satisfactorio con un porcentaje de precisión del 91% para datos nuevos. Se ha logrado diseñar una red neuronal artificial basada en el lenguaje de programación Python, capaz de ofrecer un resultado diagnóstico y probabilístico. Este logro ha figurado un avance significativo en el campo de la medicina y podría mejorar el diagnóstico y tratamiento de la DM en el futuro. Se ha sugerido la expansión de la base de datos empleada para entrenar la red neuronal, con la finalidad de mejorar sus capacidades diagnósticas y probabilísticas.

Palabras clave: Aprendizaje automático, Inteligencia artificial, Python.

ABSTRACT

Diabetes Mellitus is currently one of the most common chronic and systemic diseases, which generates great concern in the health system due to the increase of cases in recent years. Moreover, the trend is to continue to increase in the following years. Early diagnosis is required to initiate lifestyle changes in people suffering from the disease to prevent the pathology from evolving and causing complications to the patient. Artificial intelligence and machine learning have allowed doctors to have a diagnostic support method to speed up the processes and, at the same time, reduce the amount of resources used in the process. With the goal of developing a machine learning method to provide clinical and probabilistic diagnosis, an artificial neural network model has been designed in Python programming language to analyze preclinical data and expert-recommended laboratory tests. After having evaluated two other machine learning methods, the decision tree has been used as the validation method due to its performance and accuracy. The diagnosis provided by the neural network has been corroborated with the expert physician's criteria and it has shown a quite satisfactory performance with an accuracy percentage of 91% for new data. It has been possible to design an artificial neural network based on the Python programming language, capable of providing a diagnostic and probabilistic result. This achievement represents a significant advance in the field of personalized medicine and could improve the diagnosis and treatment of Diabetes Mellitus in the future. Expansion of the database used to train the neural network is suggested, with the aim of improving its diagnostic and probabilistic capabilities.

Key words: Artificial intelligence, Machine Learning, Python.

ÍNDICE DE CONTENIDO

I.	Introducción.....	1
II.	Planteamiento del problema.....	2
2.1	Precedentes del Problema	2
2.2	Definición del Problema.....	3
2.2.1	Enunciado del problema.....	3
2.2.2	Preguntas de Investigación.....	4
2.2.2.1	Pregunta General	4
2.2.2.2	Preguntas Específicas	4
2.3	Objetivos.....	4
2.3.1	Objetivo General.....	4
2.3.2	Objetivos Específicos.....	5
2.4	Justificación.....	6
2.5	Hipótesis	7
III.	Estado del arte.....	8
3.1	Análisis de la situación actual	8
3.1.1	Macroentorno.....	8
3.1.2	Microentorno	13
3.1.2.1	Latinoamérica	13
3.1.2.2	Honduras.....	17
3.2	Conceptualización	19
3.2.1	Diabetes mellitus	19
3.2.2	Tipos de diabetes	20

3.2.3	Métodos para el diagnóstico de diabetes	20
3.2.4	Inteligencia Artificial	21
3.3	Teorías de Sustento	25
3.3.1	Bases teóricas.....	25
3.3.1.1	Fundamentos de la inteligencia artificial	25
3.3.1.2	Principios de las redes neuronales artificiales.....	26
3.3.2	Metodologías aplicadas.....	34
3.4	Marco Legal	37
3.5	Matriz del Estado del Arte.....	39
IV.	Metodología	48
4.1	Enfoque	48
4.2	Variables de investigación.....	49
4.2.1	Variable dependiente.....	49
4.2.2	Variables independientes	49
4.3	Técnicas e instrumentos aplicados	50
4.3.1	Visual Studio Code.....	51
4.3.1.1	Pandas.....	51
4.3.1.2	Numpy.....	51
4.3.1.3	Seaborn.....	51
4.3.1.4	Matplotlib	52
4.3.1.5	Keras	52
4.3.1.6	Tensorflow	52
4.3.1.7	Cv2.....	53

4.3.1.8	Tkinter	53
4.3.1.9	Scikit-learn.....	54
4.3.2	Minitab.....	54
4.3.3	Microsoft Excel	54
4.4	Materiales	55
4.5	Población y muestra.....	55
4.6	Metodología de estudio.....	57
4.6.1	Fase de diseño	57
4.6.1.1	Investigación de variables asociadas.....	57
4.6.1.2	Definición de variables de entrada para la red neuronal.....	57
4.6.1.3	Selección del lenguaje de programación e IDE.....	59
4.6.1.4	Definición de arquitectura de software y métodos de ML.....	59
4.6.2	Fase de validación	59
4.6.2.1	Entrenamiento de los métodos de inteligencia artificial	59
4.6.2.2	Validación de coherencia entre resultados.....	60
4.6.2.3	Análisis de los variables predictivas y diagnósticas de la red neuronal	60
4.6.2.4	Desarrollo de interfaz gráfica	60
4.7	Metodología de validación	61
4.8	Cronograma de actividades.....	62
4.9	Matriz metodológica	63
4.10	Operacionalización de variables	65
V.	Resultados y análisis	67
5.1	Relación de variables de entrada para el sistema	67

5.2	Entrenamiento de métodos de aprendizaje automático.....	70
5.2.1	Red neuronal artificial con Keras y TensorFlow	70
5.2.2	K-Nearest Neighbor Algorithm.....	75
5.2.3	Árbol de decisión.....	75
5.3	Validación de coherencia entre resultados.....	76
5.3.1	Árbol de decisión.....	76
5.3.2	K-Nearest Neighbor	78
5.3.3	Red neuronal con Keras y TensorFlow	80
5.4	Análisis de los variables predictivas y diagnósticas.....	83
5.5	Desarrollo de interfaz gráfica.....	85
5.6	Análisis estadístico descriptivo.....	87
5.6.1	Análisis de edad	90
5.6.2	Análisis de peso.....	93
5.6.3	Análisis de glucosa en sangre.....	96
5.6.4	Análisis de hemoglobina glicosilada.....	99
5.6.5	Análisis de colesterol.....	102
5.6.6	Análisis de presión arterial sistólica.....	105
5.6.7	Análisis de presión arterial diastólica.....	108
5.6.8	Análisis de triglicéridos.....	111
5.6.9	Análisis de creatinina.....	114
VI.	Conclusiones.....	117
VII.	Recomendaciones.....	119
VIII.	Aplicabilidad/ Implementación.....	120

IX. Evolución de trabajo actual / trabajo futuro	121
X. Concordancia de los segmentos de la tesis con la investigación.....	122
Bibliografía.....	126
Anexos.....	135

ÍNDICE DE ILUSTRACIONES

Ilustración 1 Proyección del incremento de la DM 2015-2040.	9
Ilustración 2 Costes directos en España relacionados con la DM y sus complicaciones.....	10
Ilustración 3 Esquema de aprendizaje supervisado.....	22
Ilustración 4 Esquema de aprendizaje no supervisado.....	23
Ilustración 5 Relación de la inteligencia artificial con enfoques que involucran algún tipo de aprendizaje.....	24
Ilustración 6 Organización de las capas del modelo DMLP.....	27
Ilustración 7 Organización de las capas del modelo Hopfield.	28
Ilustración 8 Organización de las capas del modelo SOMs.....	28
Ilustración 9 Organización de las capas del modelo DAE.	29
Ilustración 10 Organización de las capas del modelo RNN.....	29
Ilustración 11 Organización de las capas del modelo RBM.....	30
Ilustración 12 Organización de las capas del modelo CNN.....	30
Ilustración 13 Diagrama de neurona artificial.	31
Ilustración 14 Identificación de variable dependiente y variables independientes.....	49
Ilustración 15 Metodología en V.....	61
Ilustración 16 Mapa de calor de análisis de grupos.	67

Ilustración 17 Diagrama de dispersión de valores de glucosa en sangre en pacientes sin diagnóstico de DM.....	68
Ilustración 18 Diagrama de dispersión de valores de glucosa en sangre en pacientes diabéticos.....	69
Ilustración 19 Diagrama de dispersión de valores de hemoglobina glicosilada en pacientes sin diagnóstico de DM.....	69
Ilustración 20 Diagrama de dispersión de valores de hemoglobina glicosilada en pacientes diabéticos.....	70
Ilustración 21 Función de pérdida con respecto a las épocas en la red neuronal con Keras y TensorFlow.....	71
Ilustración 22 Precisión del modelo de la red neuronal con Keras y TensorFlow.....	72
Ilustración 23 Funcionamiento lógico de la red neuronal con TF y Keras.....	74
Ilustración 24 Error en el modelo de k-NN con respecto al k utilizado.....	75
Ilustración 25 Modelo de árbol de decisión.....	76
Ilustración 26 Matriz de confusión según la clasificación realizada por el modelo de árbol de decisión.....	77
Ilustración 27 Clasificación de pacientes según el modelo kNN utilizando la base de datos de entrenamiento.....	79
Ilustración 28 Agrupación de valores de predicción por el modelo kNN utilizando valores de k diferentes.....	80
Ilustración 29 Interfaz gráfica desarrollada en Visual Studio Code.....	85
Ilustración 30 Interfaz gráfica, resultado de diagnóstico positivo.....	86
Ilustración 31 Interfaz gráfica, resultado de diagnóstico negativo.....	86
Ilustración 32 Interfaz gráfica, resultado de diagnóstico negativo probabilístico.....	87

Ilustración 33 (Histograma con curva normal) de Edad asociado a pacientes sin diagnóstico de DM.....	91
Ilustración 34 Gráfica de valores individuales de edad asociada a pacientes sin diagnóstico de DM.....	91
Ilustración 35 (Histograma con curva normal) de Edad asociado a pacientes con DM.....	92
Ilustración 36 Gráfica de valores individuales de edad asociada a pacientes con DM.	92
Ilustración (Histograma con curva normal) de Peso asociado a pacientes sin diagnóstico de DM.	94
Ilustración 38 Gráfica de valores individuales de peso asociada a pacientes sin diagnóstico de DM.....	94
Ilustración 39 (Histograma con curva normal) de Peso asociado a pacientes con DM.....	95
Ilustración 40 Gráfica de valores individuales de peso asociada a pacientes con DM.....	95
Ilustración (Histograma con curva normal) de glucosa en sangre asociado a pacientes sin diagnóstico de DM.....	97
Ilustración Gráfica de valores individuales de glucosa en sangre asociada a pacientes sin diagnóstico de DM.....	97
Ilustración (Histograma con curva normal) de glucosa en sangre asociado a pacientes con DM.	98
Ilustración Gráfica de valores individuales de glucosa en sangre asociada a pacientes con DM	98
Ilustración 45 (Histograma con curva normal) de hemoglobina glicosilada asociado a pacientes sin diagnóstico de DM.....	100
Ilustración 46 Gráfica de valores individuales de hemoglobina glicosilada asociada a pacientes sin diagnóstico de DM.....	100
Ilustración 47 (Histograma con curva normal) de hemoglobina glicosilada asociado a pacientes con DM.....	101

Ilustración 48 Gráfica de valores individuales de hemoglobina glicosilada asociada a pacientes sin con DM.....	101
Ilustración 49 (Histograma con curva normal) de colesterol asociado a pacientes sin diagnóstico de DM.....	103
Ilustración 50 Gráfica de valores individuales de colesterol asociada a pacientes sin diagnóstico de DM.....	103
Ilustración 51 (Histograma con curva normal) de colesterol asociado a pacientes con DM.	104
Ilustración 52 Gráfica de valores individuales de colesterol asociada a pacientes con DM.....	104
Ilustración 53 (Histograma con curva normal) de PAS asociado a pacientes sin diagnóstico de DM.	106
Ilustración 54 Gráfica de valores individuales de PAS asociada a pacientes sin diagnóstico de DM	106
Ilustración 55 (Histograma con curva normal) de PAS asociado a pacientes con DM.....	107
Ilustración 56 Gráfica de valores individuales de PAS asociada a pacientes con DM	107
Ilustración 57 (Histograma con curva normal) de PAD asociado a pacientes sin diagnóstico de DM.	109
Ilustración 58 Gráfica de valores individuales de PAD asociada a pacientes sin diagnóstico de DM	109
Ilustración 59 (Histograma con curva normal) de PAD asociado a pacientes con DM.....	110
Ilustración 60 Gráfica de valores individuales de PAD asociada a pacientes con DM.....	110
Ilustración 61 (Histograma con curva normal) de Triglicéridos asociado a pacientes sin diagnóstico de DM.....	112
Ilustración 62 Gráfica de valores individuales de Triglicéridos asociada a pacientes sin diagnóstico de DM	112
Ilustración 63 (Histograma con curva normal) de Triglicéridos asociado a pacientes con DM...	113

Ilustración 64 Gráfica de valores individuales de Triglicéridos asociada a pacientes con DM.....	113
Ilustración 65 (Histograma con curva normal) de Creatinina asociado a pacientes sin diagnóstico de DM.....	115
Ilustración 66 Gráfica de valores individuales de Creatinina asociada a pacientes sin diagnóstico de DM.....	115
Ilustración 67 (Histograma con curva normal) de Creatinina asociado a pacientes con DM.	116
Ilustración 68 Gráfica de valores individuales de Creatinina asociada a pacientes con DM.....	116

ÍNDICE DE TABLAS

Tabla 1 Hipótesis de investigación.....	7
Tabla 2 Prevalencia de la Diabetes Mellitus en Latinoamérica.....	14
Tabla 3 Funciones de Activación más utilizadas.	33
Tabla 4 Normativa Internacional.....	37
Tabla 5 Normativa Nacional.....	38
Tabla 6 Matriz del Estado del Arte.	39
Tabla 7 Metodología de la Investigación.....	48
Tabla 8 Instrumentos utilizados.....	50
Tabla 9 Valores de Z.....	55
Tabla 10 Identificación de variables de entrada de la red neuronal.....	58
Tabla 11 Diagrama de Gantt.....	62
Tabla 12 Matriz Metodológica.....	63
Tabla 13 Operacionalización de Variables.....	65
Tabla 14 Reporte de clasificación del modelo de árbol de decisión.....	78
Tabla 15 Validación resultados obtenidos en la red neuronal.	81

Tabla 16 Comparación de resultados obtenidos.	82
Tabla 17 Comparación de modelos de red neuronal alterando las entradas.....	83
Tabla 18 Comparación de resultados obtenidos.	84
Tabla 19 Estadística descriptiva asociada a pacientes sin diagnóstico de DM.....	87
Tabla 20 Estadística descriptiva asociada a pacientes con DM.	88
Tabla 21 Concordancia de los segmentos de la tesis con la investigación.	122

ÍNDICE DE ANEXOS

Anexo 1 Función de activación utilizada en la capa de entrada.	135
Anexo 2 Función de activación utilizada en la capa oculta.....	135
Anexo 3 Base de datos de prueba.....	136

ÍNDICE DE ECUACIONES

Ecuación 1 Fórmula para cálculo de muestra de tipo no probabilístico con población desconocida.	55
Ecuación 2 Aplicación de fórmula para cálculo de muestra.	56

LISTA DE SIGLAS

ADA	Asociación Americana de Diabetes (American Diabetes Association)
ANN	Redes Neuronales Artificiales (Artificial Neural Network)
AVAD	Años de Vida Ajustados por Discapacidad
BDT	Árbol de decisión reforzado (Boosted decision tree)
CNN	Red Neuronal Convolutacional (Convolutional Neural Network)
DAE	Codificador Automático Profundo (Deep auto-encoder)
DFT	Bosque de árboles de decisión (Decision tree forest)
DL	Aprendizaje Profundo (Deep Learning)
DM	Diabetes Mellitus
DM1	Diabetes Mellitus tipo 1
DM2	Diabetes Mellitus tipo 2
DMLP	Perceptrón Multicapa Profundo (Deep Multilayer Perceptron)
FID	Federación Internacional de Diabetes
FINDRISC	Puntaje Finlandés de Riesgo de Diabetes (Finnish Diabetes Risk Score)
HbA1c	Hemoglobina Glicosilada
IA	Inteligencia Artificial
IDE	Entorno de desarrollo integrado (Integrated Development Environment)
k-NN	k-Nearest Neighbor (k-vecinos más cercanos)
ML	Aprendizaje automático (Machine Learning)
MLP	Perceptrón Multicapa (Multilayer Perceptron)
N	Tamaño de población.

N*	Tamaño de población faltante.
NAcum	Tamaño de población acumulada.
OMS	Organización Mundial de la Salud
PAD	Presión Arterial Diastólica
PAS	Presión Arterial Sistólica
RBM	Máquina de Boltzmann Restringida (Restricted Boltzmann Machine)
RNA	Redes Neuronales Artificiales
RNN	Red Neuronal Recurrente (Recurrent Neural Network)
SESAL	Secretaría de Salud
SDT	Árbol de decisión simple (Single decision tree)
SOMs	Mapas Autoorganizados (Self-organizing Maps)
SVM	Máquina de Vectores de Soporte (Support Vector Machine)

GLOSARIO

Asimetría	Medida de la falta de simetría en una distribución de datos. En estadística, se utiliza para describir la forma de una distribución de datos y puede ser positiva, negativa o cero. (Walpole et al., 2012)
Coefficiente de variación	Medida adimensional y se utiliza comúnmente para comparar la variabilidad de diferentes conjuntos de datos que pueden tener diferentes unidades de medida. (Walpole et al., 2012)
Curtosis	Medida de la forma de una distribución de datos que indica la "apuntamiento" o "aplanamiento" de la curva de frecuencia en relación a una distribución normal. (Wackerly et al., s. f.)
Desviación estándar	Medida estadística que indica la cantidad de variación o dispersión de un conjunto de datos con respecto a su media o promedio. (Wackerly et al., s. f.)
Error estándar	Medida estadística que se utiliza para estimar la variabilidad en la media de una población a partir de una muestra de datos. (Wackerly et al., s. f.)
Hiperglucemia	Nivel de glucosa en la sangre superior al normal. (ASALE & RAE, s. f.)
Homeostasis	Autorregulación de la constancia de las propiedades de un sistema influido por agentes exteriores. (ASALE & RAE, s. f.)
Máximo	Valor más grande de un conjunto de datos. (Walpole et al., 2012)

Media aritmética	Resultado de dividir la suma de varias cantidades por el número de ellas. (Walpole et al., 2012)
Mediana	Medida de posición que indica el valor que divide un conjunto de datos ordenados en dos partes iguales. (Wackerly et al., s. f.)
Mínimo	Valor más pequeño de un conjunto de datos. (Wackerly et al., s. f.)
Moda	Medida de tendencia central en estadística que representa el valor o valores que ocurren con mayor frecuencia en un conjunto de datos. (Wackerly et al., s. f.)
Porcentaje	Medida de proporción o frecuencia relativa expresada como un valor porcentual. (Wackerly et al., s. f.)
Porcentaje acumulado	Medida que indica el porcentaje acumulado de una variable en una muestra o población. (Wackerly et al., s. f.)
Primer cuartil	Medida de posición que indica el valor en el cual el 25% de los datos de una muestra o población son iguales o menores que él. (Wackerly et al., s. f.)
Rango	Medida descriptiva que indica la diferencia entre el valor máximo y el valor mínimo en un conjunto de datos. (Wackerly et al., s. f.)
Rango intercuartil	Medida de dispersión que indica la amplitud o el rango de la mitad central de un conjunto de datos, desde el primer cuartil (Q1) hasta el tercer cuartil (Q3). (Wackerly et al., s. f.)
Screening	Test para detección precoz de una determinada enfermedad. (National Cancer Institute, 2011)

Tercer cuartil	Medida de posición en estadística que indica el valor en el cual el 75% de los datos de una muestra o población son iguales o menores que él. (Wackerly et al., s. f.)
Varianza	Se expresa en unidades cuadradas de los datos originales y proporciona una medida de cuánto los datos se desvían de la media. (Walpole et al., 2012)

I. INTRODUCCIÓN

Actualmente, las redes neuronales han generado un gran impacto en distintos campos, incluyendo la medicina. Estas permiten apoyar a los profesionales de la salud a realizar mejoras en los protocolos de diagnóstico y pronóstico de enfermedades, aumentando la velocidad y reduciendo los errores médicos. (Sarmiento-Ramos, 2020)

La diabetes mellitus es una enfermedad que afecta a la gran mayoría de la población a nivel mundial. Ávila (2022) menciona que, globalmente, se estima que 346 millones de personas tienen diabetes, y se prevé que esta cifra se duplique para el año 2030. Esta patología representa una de las principales causas de muerte y también una carga significativa para el personal de salud.

Aunque en la actualidad ya existen diversos aparatos utilizados para llevar un mejor control de la enfermedad, la tendencia es que el número de casos continúe en aumento. Por ende, es de suma importancia el diagnóstico temprano de la enfermedad y la posibilidad de determinar un porcentaje de probabilidad de padecer la patología. De esta forma, la persona puede realizar cambios en su estilo de vida para evitar complicaciones derivadas de la diabetes y tener una esperanza de vida mayor.

El proyecto de investigación se realizará con el fin de desarrollar una red neuronal que permita el diagnóstico y probabilidad de padecer diabetes de forma rápida y precisa evaluando los parámetros correspondientes recomendados por la SESAL y la ADA.

En el capítulo II se presentarán generalidades del trabajo investigativo para dar a entender el origen de la problemática. Luego, en el capítulo III se proporcionará una explicación detallada del contexto de la problemática y la inteligencia artificial.

El capítulo IV expondrá la metodología a utilizar para finalmente, exhibir en el capítulo V los resultados obtenidos. Seguido de las conclusiones y recomendaciones en el capítulo VI y VII respectivamente. Además, en el capítulo VIII se explicarán los campos de aplicación del trabajo y finalizar brindando observaciones sobre la evolución futura del trabajo en el capítulo IX.

II. PLANTEAMIENTO DEL PROBLEMA

Este capítulo brinda una breve introducción sobre el tema central de la investigación, el problema a resolver y la importancia de la misma. Adicionalmente se analizaron algunos trabajos previos realizados con un enfoque similar indicando los resultados que se obtuvieron en estos. Finalmente, se plasmaron las preguntas de investigación ligadas directamente a los objetivos que se pretenden alcanzar al finalizar el trabajo.

2.1 PRECEDENTES DEL PROBLEMA

Aproximadamente, 400 millones de individuos en todo el mundo padecen diabetes. Algunos expertos consideran esto como “un grave problema global” cuya tendencia es la de seguir en aumento. (*La diabetes a través de las estadísticas*, 2018)

En 2021 la tasa mundial de afectados rondaba los 536,600 millones según el atlas de la FDI. (Federación Internacional de Diabetes, 2021)

Durante el año 2019, la diabetes fue posicionada como la sexta causa principal de muerte en las Américas, con aproximadamente 244,084 muertes causadas directamente por dicha patología. Además, esta es la segunda causa principal de años de vida ajustados por discapacidad (AVAD), lo que refleja las complicaciones limitantes que las personas con diabetes sufren a lo largo de su vida. (Diabetes - OPS/OMS | Organización Panamericana de la Salud, 2022). A nivel mundial, entre 2000 y 2016, se observó un aumento del 5% en la mortalidad prematura por diabetes.

La problemática global de la diabetes mellitus tiene múltiples causas y factores que contribuyen a su creciente incidencia en todo el mundo. Un informe de la OMS indica una asociación entre el desarrollo de DM con el aumento de peso, la obesidad abdominal, el sedentarismo y malos hábitos alimenticios. (Carolino et al., 2008). Por lo tanto, a medida aumenta la prevalencia de la obesidad y el sedentarismo en todo el mundo, también lo hace la incidencia de la diabetes.

El envejecimiento de la población también se asocia con un mayor riesgo de padecer diabetes. En otras palabras, se puede decir que el envejecimiento demográfico y la alteración del

estilo de vida son considerados los principales determinantes del fuerte aumento de casos de la DM2 en los últimos años. (Carolino et al., 2008)

Por último, otro factor que también puede incidir de forma negativa es la falta de acceso a la atención médica y los tratamientos adecuados, esto contribuye a la problemática global de la diabetes, especialmente en países de bajos y medianos ingresos.

2.2 DEFINICIÓN DEL PROBLEMA

La DM muchas veces pasa desapercibida debido a la ausencia de síntomas en los inicios de la enfermedad. Dentro de los métodos diagnósticos utilizados actualmente se encuentra la glucemia plasmática mediante análisis de laboratorio y la glucemia capilar utilizando dispositivos conocidos como glucómetros. Aunque el primero tiene un gran impacto a la hora de diagnosticar la patología, con ayuda de la inteligencia artificial se puede brindar un apoyo al personal de salud agilizando los procesos con el mínimo porcentaje de error. Por otro lado, la glucemia capilar es el método diagnóstico más sobresaliente debido a la rapidez en la obtención del resultado, sin embargo, este también tiene sus limitaciones que pueden afectar la validez de la lectura. En su mayoría, estas corresponden a errores operacionales en el procedimiento de la toma de la muestra capilar como ser la inadecuada cantidad de sangre, tiempo de lectura inadecuado, sangre aplicada incorrectamente, inserción incorrecta de la tira o uso de tiras caducadas. (Pariente Rodrigo et al., 2017)

2.2.1 ENUNCIADO DEL PROBLEMA

El estudio pretende desarrollar un modelo de aprendizaje automático capaz de analizar grandes cantidades de datos clínicos de pacientes del Hospital El Progreso y utilizarlos para generar diagnósticos precisos y probabilísticos que puedan ser utilizados para mejorar la atención de los pacientes con Diabetes Mellitus. El desafío radica en identificar las variables clínicas relevantes e implementar un modelo de inteligencia artificial adecuado para realizar el diagnóstico de manera eficaz y precisa.

2.2.2 PREGUNTAS DE INVESTIGACIÓN

2.2.2.1 PREGUNTA GENERAL

¿Cómo desarrollar un método de diagnóstico para la Diabetes Mellitus con capacidades clínicas probabilísticas utilizando inteligencia artificial en El Progreso, Honduras, en los meses de enero a marzo del 2023?

2.2.2.2 PREGUNTAS ESPECÍFICAS

1. ¿Cuál es el enfoque de aplicabilidad que se le ha dado al aprendizaje automático en el contexto clínico diagnóstico de la Diabetes Mellitus por parte de otros investigadores en el periodo 2013-2023?
2. ¿Cuál es la funcionalidad y aplicabilidad de distintos métodos de inteligencia artificial como herramientas diagnósticas y probabilísticas de la Diabetes Mellitus en pacientes del Hospital El Progreso, Honduras, en los meses de enero a marzo del 2023?
3. ¿Cuáles son los principios computacionales y matemáticos que se pueden aplicar en el desarrollo de una red neuronal alimentada con datos de 170 pacientes del Hospital El Progreso, Honduras, en los meses de enero a marzo del 2023?
4. ¿Qué efecto tiene el uso de librerías durante el desarrollo de una red neuronal para diagnosticar la Diabetes Mellitus y determinar la probabilidad de padecer la patología en pacientes del Hospital El Progreso, Honduras, en los meses de enero a marzo del 2023?

2.3 OBJETIVOS

2.3.1 OBJETIVO GENERAL

Desarrollar un método de diagnóstico para la Diabetes Mellitus con capacidades clínicas probabilísticas utilizando inteligencia artificial a través de un algoritmo computacional desarrollado en lenguaje de programación Python en El Progreso, Honduras, en los meses de enero a marzo del 2023.

2.3.2 OBJETIVOS ESPECÍFICOS

1. Recopilar información sobre aplicación del aprendizaje automático por parte de otros investigadores en el contexto clínico diagnóstico de la Diabetes Mellitus desde el año 2013 hasta la actualidad en macro y microentorno.
2. Evaluar la funcionalidad y aplicabilidad de tres métodos de inteligencia artificial como herramientas diagnósticas y probabilísticas de la Diabetes Mellitus en 170 pacientes del hospital El Progreso, Honduras, en los meses de enero a marzo del 2023.
3. Identificar los principios computacionales y matemáticos que se pueden aplicar en el desarrollo de una red neuronal realizada en los meses de enero a marzo del 2023 y alimentada con una muestra de datos de 170 pacientes hondureños.
4. Determinar el efecto que tiene el uso de librerías en el desarrollo de una red neuronal para diagnosticar la Diabetes Mellitus y determinar la probabilidad de padecer la patología en los meses de enero a marzo del 2023.

2.4 JUSTIFICACIÓN

La DM es una enfermedad que va en incremento con el paso de los años, comprometiendo la calidad de vida de las personas que la padecen. En Honduras entre el 5% y el 12% de la población padece la enfermedad y se estima que por cada paciente diagnosticado hay al menos uno sin diagnosticar. En 2015, se creó el Instituto Nacional del Diabético para otorgar a los pacientes diabéticos un tratamiento personalizado, con calidad y calidez en el país. (Atención a los pacientes diabéticos y a personas con problemas de salud mental en tiempos de COVID-19 en Honduras - OPS/OMS | Organización Panamericana de la Salud, s. f.).

Debido a la alta demanda de pacientes, en septiembre del 2022, se requirió extender el horario de atención desde las 7 de la mañana hasta las 7 de la noche. (Atención a los pacientes diabéticos y a personas con problemas de salud mental en tiempos de COVID-19 en Honduras - OPS/OMS | Organización Panamericana de la Salud, s. f.). Con esto podemos concluir, que hubo un notable incremento en el número de casos y, posiblemente, también un aumento en los recursos utilizados para diagnosticar y tratar la enfermedad.

En los inicios de la patología no siempre se presentan síntomas que sean indicios de padecerla, en consecuencia, la DM se conoce como una enfermedad silenciosa. Por este motivo, existe gran parte de la población que vive con la enfermedad y no se da cuenta hasta que comienzan a surgir complicaciones asociadas a dicha patología.

El presente proyecto proporcionará un prototipo de red neuronal para diagnosticar y brindar un porcentaje de probabilidad de padecer DM. Esta red será entrenada con datos obtenidos de expedientes clínicos de pacientes del Hospital El Progreso con el fin de que la red, luego del entrenamiento, muestre resultados certeros con un porcentaje de error menor al 15%.

Esta investigación centra su objetivo en proporcionar un método de diagnóstico confiable de la DM al sector salud en Honduras, disminuyendo considerablemente los recursos necesarios para el proceso. Así mismo, se persigue la posibilidad de brindar un diagnóstico probabilístico de la enfermedad, es decir, determinar la posibilidad de que la persona padezca DM a través de un análisis profundo desarrollado por la red neuronal artificial.

2.5 HIPÓTESIS

Esta sección proporciona las hipótesis analíticas de la investigación vinculadas directamente con las preguntas de investigación. Estableciendo para cada pregunta de investigación una hipótesis de investigación y una hipótesis nula (Tabla 1).

Tabla 1 Hipótesis de investigación.

Preguntas de Investigación	Hipótesis
¿Cuál es el enfoque de aplicabilidad que se le ha dado al aprendizaje automático en el contexto clínico diagnóstico de la Diabetes Mellitus por parte de otros investigadores en el periodo 2013-2023?	<p>Hi: A partir del año 2013, el aprendizaje automático en el contexto clínico diagnóstico de la Diabetes Mellitus ha sido aplicable por parte de otros investigadores.</p> <p>Ho: El aprendizaje automático en el contexto clínico diagnóstico de la Diabetes Mellitus no ha sido aplicable por parte de otros investigadores a partir del año 2013.</p>
¿Cuál es la funcionalidad y aplicabilidad de distintos métodos de inteligencia artificial como herramientas diagnósticas y probabilísticas de la Diabetes Mellitus en pacientes del Hospital El Progreso, Honduras?	<p>Hi: Los métodos de inteligencia artificial son herramientas eficaces y útiles para el diagnóstico y la probabilidad de la Diabetes Mellitus.</p> <p>Ho: Los métodos de inteligencia artificial no son útiles ni eficaces para el diagnóstico y probabilidad de Diabetes Mellitus.</p>
¿Cuáles son los principios computacionales y matemáticos que se pueden aplicar en el desarrollo de una red neuronal alimentada con datos de 170 pacientes del Hospital El Progreso, Honduras en los meses de enero a marzo del 2023?	<p>Hi: Existen diversos principios matemáticos y computaciones que se pueden aplicar en el desarrollo de una red neuronal.</p> <p>Ho: No existen diversos principios matemáticos y computaciones que se pueden aplicar en el desarrollo de una red neuronal.</p>
¿Qué efecto tiene el uso de librerías durante el desarrollo de una red neuronal para diagnosticar la Diabetes Mellitus y determinar la probabilidad de padecer la patología en pacientes del Hospital El Progreso, Honduras en los meses de enero a marzo del 2023?	<p>Hi: El uso de las librerías facilita el desarrollo de una RNA para brindar un diagnóstico preciso de Diabetes Mellitus e identificar la probabilidad de padecer la patología.</p> <p>Ho: No facilita el proceso de desarrollo la implementación de librerías durante la creación de una RNA para brindar un diagnóstico preciso de Diabetes Mellitus e identificar la probabilidad de padecer la patología.</p>

Fuente: Elaboración propia.

Siendo:

Hi: Hipótesis de investigación.

Ho: Hipótesis nula.

III. ESTADO DEL ARTE

En este capítulo se observó la problemática realizando un análisis de la situación actual tomando dos panoramas, el macroentorno y el microentorno. Así mismo, con el propósito de contextualizar al lector se recopiló información relevante sobre conceptos necesarios para el entendimiento total del proyecto, además de metodologías aplicadas por otros autores con investigaciones comparables a la presente.

3.1 ANÁLISIS DE LA SITUACIÓN ACTUAL

Las redes neuronales artificiales y otros métodos de inteligencia artificial comienzan a adentrarse cada vez más en la rama de la medicina en numerosos países alrededor del mundo. Se les considera una herramienta extremadamente útil para apoyar al personal de salud en la toma de decisiones en caso de dificultad para la utilización de pruebas adicionales por su costo o disponibilidad.

En esta sección se determina la situación actual enlazada directamente a los costos derivados de la problemática además de técnicas predictivas y de diagnóstico implementando redes neuronales desde la perspectiva del macroentorno y microentorno.

3.1.1 MACROENTORNO

La diabetes es una de las enfermedades más prevalentes a nivel mundial. Diversos investigadores han realizado estudios epidemiológicos para estimar la cantidad de personas que padecen la enfermedad y calcular una tendencia o proyección de la carga a largo plazo. Según Herman (2017), uno de los estudios estimaba que, entre 2010 y 2030, la cantidad de adultos con diabetes aumentaría de forma global un 54% aproximadamente.

La Federación Internacional de Diabetes demostró en el año 2015 que el top 10 de las naciones con mayor cantidad de personas diabéticas eran:

- China con 109.6 millones
- India con 99.2 millones
- Estados Unidos con 29.3 millones
- Brasil con 14.3 millones

- Rusia con 12.1 millones
- México con 11.5 millones
- Indonesia con 10 millones
- Egipto con 7.8 millones
- Japón con 7.2 millones
- Bangladesh con 7.1 millones

(Dagogo-Jack, 2017)

En referencia al ranking anterior, cabe resaltar que en el año 1980 se tenía a Alemania, Ucrania, Italia y Reino Unido dentro de los 10 países con el mayor número de personas con diabetes. Sin embargo, debido a un incremento bastante abrupto en el número de casos en Indonesia, Pakistán, México y Egipto, estos quedaron excluidos del ranking para el año 2014.

Por otro lado, gracias a iniciativas nacionales de prevención de la diabetes, la cantidad de personas con diabetes no incrementará drásticamente en Estados Unidos. Debido a esto, no hay un gran cambio en la proyección del incremento de la DM en Norte América para el año 2040 (Ilustración 1).

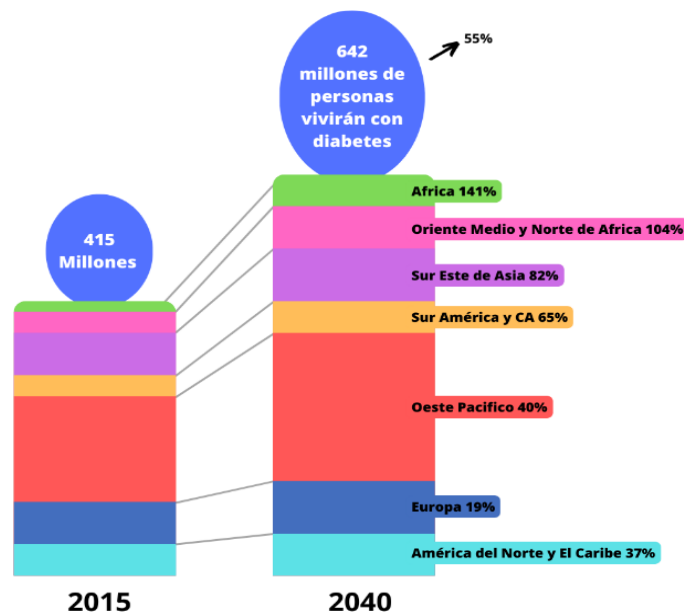


Ilustración 1 Proyección del incremento de la DM 2015-2040.

Fuente: (Dagogo-Jack, 2017)

Así mismo, esta enfermedad atribuye a una elevada inversión monetaria distribuida en costes hospitalarios, atención ambulatoria y atención primaria. Incluyendo pruebas complementarias y tiras reactivas de auto monitorización de glucosa capilar y fármacos. (Pariente Rodrigo et al., 2017). Después de lo anterior expuesto, se pueden apreciar los costes directos de España que, sin ser uno de los países con la mayor cantidad de casos, siempre se observa una elevada suma de gastos (Ilustración 2).

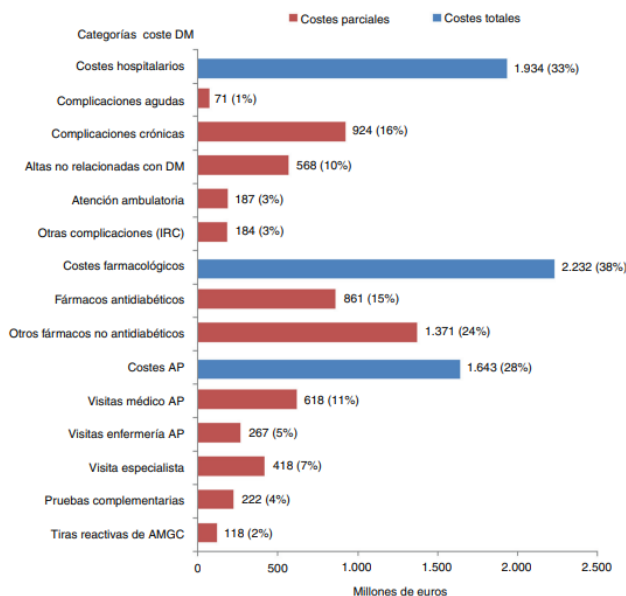


Ilustración 2 Costes directos en España relacionados con la DM y sus complicaciones.

Fuente: (Crespo et al., 2013)

En neurología, se han utilizado las redes neuronales con un enfoque orientado a la detección, clasificación y predicción mediante el análisis del electroencefalograma. (Delgado et al., 2019). Esto ha permitido la detección automática de crisis epilépticas además de su clasificación a través de redes neuronales entrenadas para el análisis de 14 características del electroencefalograma para determinar el estado del cerebro.

Así mismo, se han utilizado las redes neuronales artificiales para el diagnóstico del dengue (Ríos, 2022), la toma de decisiones clínicas en cuanto al tratamiento de insuficiencia renal (Martínez Bernabé et al., 2014) y, recientemente, para determinar la extensión y gravedad de la afección pulmonar causada por COVID-19 mediante el análisis de tomografías computarizadas de tórax (Carvalho et al., 2020).

Además, la aplicación de la inteligencia artificial y las redes neuronales se han considerado de gran apoyo para el soporte de decisiones médicas en estudiantes de medicina.

En cuanto al uso de aprendizaje automático aplicado a la diabetes mellitus, en el año 2020 se realizó en India una revisión bibliográfica titulada "Machine learning and artificial intelligence based Diabetes Mellitus detection and self-management: A systematic review". Esta se trata de una revisión sistemática de estudios que han utilizado técnicas de aprendizaje automático e inteligencia artificial para detectar y autogestionar la diabetes mellitus. (Chaki et al., 2020)

En la revisión, los autores examinaron la literatura existente sobre el uso de estas tecnologías para la detección y autogestión de la diabetes mellitus y encontraron un total de 28 estudios relevantes. Estos estudios incluyeron diferentes enfoques de aprendizaje automático y técnicas de inteligencia artificial, como redes neuronales artificiales, árboles de decisión y algoritmos genéticos.

Los resultados de la revisión indicaron que el aprendizaje automático y la inteligencia artificial pueden ser muy efectivos en la detección y autogestión de la diabetes mellitus. Se encontró que estas tecnologías pueden utilizarse para predecir el riesgo de diabetes mellitus en pacientes, monitorear los niveles de glucosa en sangre, proporcionar recomendaciones personalizadas de dieta y ayudar en la autogestión de la enfermedad. (Chaki et al., 2020)

No obstante, la revisión también identificó varios desafíos y limitaciones en el uso de tecnologías de aprendizaje automático e inteligencia artificial para la detección y autogestión de la diabetes mellitus. Entre ellos, se encuentra la necesidad de contar con datos de alta calidad y la dificultad de integrar estas tecnologías en la práctica clínica.

En general, este estudio destaca el potencial de las tecnologías de aprendizaje automático e inteligencia artificial para mejorar la detección y autogestión de la diabetes mellitus, pero también señala la necesidad de abordar los desafíos y limitaciones para lograr una implementación exitosa y efectiva de estas tecnologías en la práctica clínica. (Chaki et al., 2020)

En Irán se desarrolló un estudio titulado "Application of an artificial neural network model for diagnosing type 2 diabetes mellitus and determining the relative importance of risk factors".

En él se empleó una red neuronal artificial para diagnosticar la diabetes tipo 2 y evaluar la importancia relativa de los factores de riesgo asociados. (Borzouei & Soltanian, 2018)

En la investigación, se utilizaron datos de pacientes con diabetes tipo 2 y pacientes sanos para entrenar la red neuronal artificial en el diagnóstico de la enfermedad. Posteriormente, se utilizó la red neuronal para evaluar la importancia relativa de varios factores de riesgo asociados con la diabetes tipo 2, tales como la edad, el índice de masa corporal y el nivel de actividad física.

Los resultados obtenidos demostraron que la red neuronal fue capaz de diagnosticar la diabetes tipo 2 con una precisión superior al 90%. Asimismo, el análisis de los factores de riesgo reveló que la edad y el índice de masa corporal eran los factores más importantes en la predicción de la diabetes tipo 2. (Borzouei & Soltanian, 2018)

Este estudio es relevante para la detección temprana y prevención de la diabetes tipo 2. Mediante la utilización de una red neuronal artificial, los investigadores lograron identificar los factores de riesgo más importantes relacionados con la enfermedad, lo cual podría permitir una detección temprana y una intervención preventiva.

El artículo "Continuous blood glucose level prediction of Type 1 Diabetes based on Artificial Neural Network" desarrollado en Francia en 2018 se centra en un estudio que utiliza una red neuronal artificial para predecir de manera continua los niveles de glucosa en sangre en pacientes con diabetes tipo 1. (Ben Ali et al., 2018)

Los investigadores utilizaron datos de pacientes con diabetes tipo 1 para entrenar la red neuronal artificial en la predicción de los niveles de glucosa en sangre. Posteriormente, la red neuronal se utilizó para predecir de manera continua los niveles de glucosa en sangre durante un período de tiempo determinado.

Los resultados del estudio mostraron que la red neuronal artificial fue capaz de predecir de manera precisa y continua los niveles de glucosa en sangre en pacientes con diabetes tipo 1. Además, la red neuronal también fue capaz de detectar patrones y tendencias en los niveles de

glucosa en sangre, lo que podría ayudar a los pacientes a ajustar su tratamiento y mejorar su control glucémico.

Este estudio es significativo ya que la predicción continua de los niveles de glucosa en sangre es una herramienta valiosa para el manejo de la diabetes tipo 1. La capacidad de detectar patrones y tendencias en los niveles de glucosa en sangre podría ayudar a los pacientes y sus médicos a tomar decisiones informadas sobre el tratamiento, reducir el riesgo de complicaciones de la diabetes y mejorar la calidad de vida de los pacientes. (Ben Ali et al., 2018)

3.1.2 MICROENTORNO

Evaluando la situación desde la perspectiva del microentorno, se recopiló información relevante sobre la DM en Latinoamérica. Haciendo hincapié en la región Centroamericana, sobre todo en la situación de Honduras con respecto a esta problemática.

3.1.2.1 LATINOAMÉRICA

Dos de los diez países líderes en el ranking de diabetes a nivel mundial expuesto anteriormente son parte de Latinoamérica, Brasil (14,3 millones) y México (11,5 millones). Además, se espera que el incremento en el número de casos sea mayor en esta región en comparación con otras áreas. (Dagogo-Jack, 2017).

Así mismo, según la OMS, la prevalencia de la obesidad en esta región ha ido aumentando drásticamente con el paso de los años. Siendo este uno de los factores de riesgo de la diabetes, significa entonces que la población en Latinoamérica tiene más probabilidades de padecer la patología.

Como puede observarse, Brasil y México son los 2 países de Latinoamérica con mayor cantidad de casos de DM (Tabla 2). De igual manera, estos tienen el mayor número de muertes debidas a diabetes y Brasil posee el costo anual de tratamiento más elevado de todos (1,031.44 USD).

Tabla 2 Prevalencia de la Diabetes Mellitus en Latinoamérica.

País	Número de casos	Prevalencia	Tasa de muerte anual debido a diabetes	Costo anual del tratamiento por sujeto (USD)	Número estimado de casos sin diagnosticar	Incremento anual del número de casos
Brasil	14,300,000	10.52	129,226	1,031.44	6,164,590	37,000
México	11,500,000	14.4	80,000	815.53	3,452,410	323,000
Colombia	2,067,870	7.26	14,602	482.72	951,220	95,000
Venezuela	1,764,900	10.39	13,380	914.01	811,850	61,000
Argentina	1,570,200	5.57	15,416	966.44	722,290	29,000
Chile	1,442,610	12.78	10,459	992.13	755,600	28,000
Perú	1,108,610	6.81	8,150	307.31	509,960	-
Cuba	872,950	8.58	7,560	823.71	401,560	19,000
Guatemala	589,140	9.93	7,202	311.52	271,010	27,000
Ecuador	563,840	6.89	5,492	335.41	259,360	19,000
República Dominicana	405,508	7.36	5,183	419.28	186,570	-
Puerto Rico	391,870	12.98	-	-	108,590	-
Bolivia	325,220	6.89	4,732	124.63	149,600	-
El Salvador	312,430	9.88	3,233	333.58	143.72	-
Nicaragua	309,320	11.58	3,001	172.21	142,290	-
Costa Rica	259,350	8.81	1,659	937.20	119,300	-
Honduras	239,590	7.16	2,338	209.40	110,210	-
Paraguay	222,220	6.81	2,174	283.14	102,220	-
Panamá	184,580	8.59	1,399	732.45	84,910	-
Uruguay	157,330	6.02	1,122	922.68	72,370	-
Guayana Francesa	12,610	9.60	-	-	5,800	-

Fuente: (Dagogo-Jack, 2017)

Después de las consideraciones anteriores, se puede decir que, probablemente, la causa del gran número de muertes debidas a DM sea debido al alto costo del tratamiento o a que la población no está diagnosticada.

Por otro lado, en los países centroamericanos se encuentra Guatemala con el mayor número de casos (589,140), seguido de El Salvador (312,430), Nicaragua (309,320), Costa Rica (259,350), Honduras (239,590) y Panamá (184,580).

Actualmente, en distintas partes del mundo, la inteligencia artificial y las redes neuronales destacan en diversas ramas de la medicina, sobre todo aquellas destinadas al diagnóstico y clasificación de enfermedades. En cardiología, se ha utilizado en el diagnóstico del infarto agudo

de miocardio (Sprockel et al., 2014), En este estudio realizado en Colombia se examina el uso de redes neuronales en el diagnóstico de una enfermedad cuyo diagnóstico sigue siendo un reto a pesar de los avances médicos. El estudio utilizó datos clínicos de pacientes con sospecha de infarto agudo de miocardio para entrenar una red neuronal en la predicción del diagnóstico.

Los resultados mostraron que la red neuronal tuvo una precisión del 96,4% en la predicción del diagnóstico, superando los métodos de diagnóstico convencionales. Además, la red neuronal también identificó las variables clínicas más relevantes para el diagnóstico. El estudio sugiere que el uso de redes neuronales puede mejorar la precisión del diagnóstico de infarto agudo de miocardio y ayudar a los médicos a tomar decisiones informadas sobre el tratamiento de los pacientes, aunque se necesitan más estudios para validar estos hallazgos y evaluar la viabilidad del uso de redes neuronales en la práctica clínica.

En Perú, se ha implementado la inteligencia artificial para optimizar el diagnóstico de la tuberculosis a través de una aplicación en el dispositivo móvil del personal capacitado. El proceso se basa en tomar una fotografía de la radiografía de tórax del paciente, esta es transmitida a la aplicación diseñada en base a inteligencia artificial con redes neuronales convoluciones. Esta detecta anomalías pulmonares y finalmente sugiere manifestaciones clínicas preliminares de la tuberculosis. (Curioso et al., 2020)

En 2014 se desarrolló el artículo "Artificial Intelligence to Assist Clinical Diagnosis in Medicine" en una revista Mexicana donde se explora el potencial de la inteligencia artificial (IA) en la mejora del diagnóstico clínico en medicina. Los autores discuten cómo la IA puede ayudar a los profesionales médicos a analizar grandes cantidades de datos clínicos y proporcionar diagnósticos más precisos y eficaces. (Lugo-Reyes et al., 2014)

El artículo aborda varios temas relevantes, incluyendo la definición y tipos de IA utilizados en medicina, los desafíos y oportunidades asociados con la implementación de la IA en la práctica clínica, y las implicaciones éticas y legales de su uso. Los autores también discuten algunos ejemplos de aplicaciones de IA en la medicina, como la detección temprana de enfermedades, el seguimiento de pacientes, y la identificación de patrones en grandes conjuntos de datos.

Uno de los puntos más destacados del artículo es que los autores enfatizan la necesidad de una colaboración interdisciplinaria entre los profesionales médicos y los expertos en IA para garantizar un uso ético y seguro de la tecnología en la práctica clínica. Además, destacan la importancia de abordar la interpretación y explicación de los resultados generados por la IA para que sean comprensibles para los médicos y los pacientes. (Lugo-Reyes et al., 2014)

En resumen, el artículo es una contribución importante para la comprensión del papel que la IA puede desempeñar en el diagnóstico clínico en medicina y los desafíos que deben superarse para su implementación exitosa. Los autores proporcionan una visión general de los aspectos relevantes de la IA en medicina y destacan la necesidad de una colaboración interdisciplinaria y una atención cuidadosa a las implicaciones éticas y legales de su uso. (Lugo-Reyes et al., 2014)

En Paraguay en 2019 se desarrolló un prototipo de red utilizando la plataforma WEKA ((Waikato Environment for Knowledge Analysis) para la clasificación de la enfermedades endémicas (Diagnóstico de Enfermedades Endémicas con redes neuronales artificiales a través de WEKA). La propuesta tecnológica se basa en un sistema que utiliza redes neuronales artificiales. Este modelo consta de 4 neuronas/síntomas para la capa de entrada, 4 capas ocultas con 4 neuronas cada una, y 4 capas de salida que corresponden a enfermedades endémicas. La aplicación de redes neuronales artificiales puede brindar una clasificación óptima de enfermedades endémicas a los profesionales médicos que trabajan en hospitales públicos. Asimismo, esta tecnología puede utilizarse para clasificar otros tipos de patologías clínicas, y sobre todo para la toma de decisiones gerenciales, directivas y epidemiológicas. Los resultados de los experimentos demuestran que la tendencia es siempre la misma, gracias al alto nivel de efectividad en el procesamiento de datos. (Sánchez & Núñez, 2019)

En Costa Rica se desarrolló un estudio para predecir la incidencia de dengue en Costa Rica mediante el uso de un modelo de árbol de decisión basado en variables climáticas y socioeconómicas. Los investigadores recolectaron datos sobre casos de dengue, temperatura, precipitación, humedad y variables socioeconómicas de 81 distritos en Costa Rica. Luego utilizaron análisis de árbol de decisión para construir un modelo que pudiera predecir la

incidencia de dengue en cada distrito. Los resultados mostraron que la temperatura, la precipitación y las variables socioeconómicas eran importantes para predecir la incidencia de dengue en Costa Rica. (Murillo & Soto, 2021)

3.1.2.2 HONDURAS

Según Jonkman et al. (2020), la prevalencia de diabetes en Honduras ha aumentado a casi el doble en los últimos 30 años, con más de 2000 muertes en el año 2016 atribuidas a la diabetes y glucemia alta. De acuerdo con los datos obtenidos de la OMS, el 80% de las muertes por complicaciones de la diabetes ocurren en países de ingresos bajos y medios, como el caso Honduras. Además, se cree que el 50% de las personas que padecen DM no han sido diagnosticadas, sin embargo, este porcentaje aumenta a un 67% en países de escasos recursos sanitarios.

En parte, lo anterior se debe a que la detección de la enfermedad se realiza mediante la realización de exámenes o test que no siempre están disponibles en los centros de atención, especialmente en aquellos de países de bajos ingresos. (Castrillón et al., 2017). Posiblemente, otra razón sea por la falta de acceso a servicios de salud de parte de la población provocada por la misma ausencia de recursos.

En cuanto a la predicción de la patología, en Honduras, al igual que en otros países, se utiliza el test FINDRISC o adaptaciones del mismo. Este test fue desarrollado en Finlandia y se basa en un análisis a partir de características como ser la edad, el consumo de frutas y verduras, la actividad física que realiza, si consume algún medicamento para la hipertensión, el índice de masa corporal y la historia familiar o antecedentes familiares con respecto a la patología. (Campos & Palomino, 2018)

Esta herramienta ha demostrado ser bastante precisa en cuanto al screening de riesgo a desarrollar DM en los siguientes 10 años al igual que el "ADA risk test".

Hablando de RNA, actualmente se utiliza la IA en distintas áreas como ser el estudio desarrollado en Roatán (Computer vision neural network using YOLOv4 for underwater fish video detection In Roatan, Honduras). El artículo trata sobre el uso de una red neuronal de visión por computadora que utiliza YOLOv4 para detectar peces en videos subacuáticos en

Roatán, Honduras. La tecnología es capaz de detectar peces en tiempo real y generar estadísticas sobre la cantidad de peces presentes en el área, lo que puede ser útil para la investigación de poblaciones de peces y la conservación marina. Además, se describe el proceso de entrenamiento de la red neuronal y se discuten los resultados obtenidos en términos de precisión y eficiencia de detección. (Jimenez Nixon, 2021)

Así mismo en el área de agricultura se desarrolló una revisión bibliográfica sobre la implementación de la inteligencia artificial en el sector agrícola de Honduras. Se exploran diferentes áreas donde la IA puede ser útil en la agricultura, incluyendo la predicción del rendimiento de los cultivos, el monitoreo del clima y el suelo, la gestión de plagas y la automatización de la maquinaria agrícola. Además, se analizan las barreras que impiden la adopción de la IA en la agricultura hondureña, como la falta de infraestructura tecnológica y la escasez de recursos financieros y humanos. Se concluye que la implementación de la IA en la agricultura hondureña tiene un gran potencial para mejorar la productividad y la sostenibilidad del sector, pero requiere un esfuerzo coordinado por parte del gobierno, las empresas y los agricultores. (Perdomo et al., 2022)

Así mismo en 2022 se utilizaron algoritmos de visión computarizada para identificar enfermedades y decoloraciones en los arrecifes de coral. Los arrecifes de coral son ecosistemas importantes, pero están en peligro debido al cambio climático y la contaminación. Los investigadores han desarrollado un sistema que utiliza imágenes de drones y cámaras submarinas para detectar signos de enfermedad y decoloración en los corales. Los algoritmos de visión computarizada pueden identificar patrones y diferencias sutiles en los colores y texturas de los corales, lo que puede ayudar a los científicos a detectar problemas en los arrecifes antes de que se vuelvan graves. Este enfoque podría ayudar a mejorar la eficacia de los programas de monitoreo y conservación de arrecifes de coral y proporcionar información importante para la toma de decisiones sobre la gestión de los recursos marinos. (Bautista et al., 2022)

3.2 CONCEPTUALIZACIÓN

Los conceptos necesarios para comprender el tema de investigación son descritos a continuación. Primero se describen los conceptos relacionados a la patología, sus diversos tipos y los métodos diagnósticos actuales, contextualizando un poco el panorama actual. Sumado a esto, se exponen los términos relevantes relacionados a la inteligencia artificial, redes neuronales, algoritmo de árbol de decisión y k-nearest neighbor ya que estos presentan mayor complejidad y son relativamente nuevos y necesarios para comprender su labor dentro del entorno salud.

3.2.1 DIABETES MELLITUS

Diabetes Mellitus: Es una enfermedad metabólica crónica de la cual aún no existe cura, por lo tanto, esta solo puede ser controlada o prevenida. "Se caracteriza por generar una disminución en la capacidad del individuo para controlar los niveles de glucosa en el torrente sanguíneo" (Kaul et al., 2013). Esto se debe a que el paciente con diabetes posee un déficit total o parcial en la producción o labor de la insulina.

La insulina y el glucagón son hormonas liberadas por el páncreas que tienen una importante labor en la homeostasis de la glucosa. Mientras el glucagón aumenta los niveles de glucosa en sangre, la insulina los disminuye. (Röder et al., 2016). Por lo tanto, debe existir un equilibrio entre la secreción de ambas para mantener los niveles de glucosa dentro del rango normal.

Esta enfermedad se define como un estado de hiperglucemia en ayunas o posprandial, es decir, antes de la primera comida del día o después de haber ingerido alimentos. Y se asocia con daño, disfunción e insuficiencia de órganos diana en órganos y tejidos, entre ellos la retina, los riñones, los nervios, el corazón y los vasos sanguíneos. (Alam et al., 2014). Por tal razón, es crucial que la enfermedad se diagnostique a tiempo para prevenir que llegue a extremos donde llega a perjudicar la salud a gran escala.

3.2.2 TIPOS DE DIABETES

Según Poveda et al. (2020), existen tres tipos de diabetes principales: diabetes mellitus tipo 1, diabetes mellitus tipo 2 y diabetes gestacional. Cada uno de estos tipos de diabetes con una fisiopatología parcialmente diferente y de los cuales, la más común es la DM tipo 2.

Diabetes Mellitus tipo 1: La fisiopatología detrás de la DM1 se basa en la autodestrucción inmunológica de las células beta del páncreas, lo que produce una deficiencia absoluta o casi absoluta de la insulina. (ElSayed et al., 2022)

Diabetes Mellitus tipo 2: La DM2 es una de las enfermedades crónicas con mayor impacto en la calidad de vida de las personas, ya que pertenece al grupo de patologías que como resultado generan incapacidad física. Esta se origina debido a una combinación de resistencia a la acción de la insulina, que generalmente está asociada con la obesidad, y una inadecuada respuesta secretora compensatoria. (Enderica et al., 2019)

Diabetes Mellitus gestacional: Este tipo de diabetes surge durante el periodo de embarazo de la mujer, haciendo énfasis en que, antes del embarazo esta no padecía la enfermedad. De igual manera, consiste en niveles elevados de glucosa en la sangre con la peculiaridad de ser considerado un trastorno de la tolerabilidad a los hidratos de carbono que conduce a la hiperglucemia. (Bougherara et al., 2018). Puede causar repercusiones a futuro tanto en la salud de la madre, como en la del feto, neonato, joven o adulto.

3.2.3 MÉTODOS PARA EL DIAGNÓSTICO DE DIABETES

Glucemia: La glucemia, también llamada glicemia, hace referencia a la concentración de glucosa en la sangre. Los valores normales se encuentran entre los 4.0 y 7.0 mmol/L, o bien 82–110 mg/dL. (Gellman, 2020). Cuando se tienen concentraciones menores a los valores mencionados anteriormente se conoce como estado de hipoglucemia y, cuando se supera ese límite se considera estado de hiperglucemia. Para llevar un control de los niveles de glucosa se realiza la monitorización a través de dispositivos conocidos como glucómetros, de no ser así se pueden realizar análisis de laboratorio de forma periódica.

Glucometría: La glucometría hace referencia a un análisis rutinario esencial para las personas diabéticas.

El glucómetro es un dispositivo médico que se utiliza para determinar el nivel aproximado de glucosa en sangre. Es utilizado para controlar los niveles de glucosa en el hogar o en un centro de atención médica. El proceso para obtener una medición se basa en recolectar cierta cantidad de sangre capilar utilizando un dispositivo de punción en la yema del dedo. Esta sangre es colocada en una tira reactiva desechable y en cuestión de segundos se obtiene una lectura. (Gellman, 2020)

Estas mediciones permiten que los profesionales de salud tomen decisiones médicamente relevantes sobre el tratamiento de los pacientes. (Schnell et al., 2015). Así también, proporciona a los pacientes una manera de llevar el control de sus niveles de glucosa desde su hogar para identificar cuando deba acudir al médico o hacer cambios en el estilo de vida.

Análisis de laboratorio: Las pruebas de laboratorio se utilizan de forma rutinaria para el diagnóstico de la diabetes. La mayoría de los fenotipos incluyen el concepto de glucosa anormal, típicamente definida como glucosa aleatoria mayor o igual a 200 mg/dl, glucosa en ayunas mayor o igual a 126 mg/dl, o un resultado de laboratorio de hemoglobina (HbA1c) de 6.5 % o más. (Richesson et al., 2013).

3.2.4 INTELIGENCIA ARTIFICIAL

Inteligencia artificial: La IA se define como una de las ramas de la informática que se enfoca en modelar las capacidades y la agudeza del cerebro humano. Así mismo, es concebida como parte de las ciencias de la computación que permite brindar diversos métodos, técnicas y herramientas para el modelado y resolución de problemas simulando el comportamiento de los seres cognitivos. Teniendo en cuenta también sus límites y posibilidades, caracterizándola como un desafío de gran complejidad. (Ocaña-Fernández et al., 2019)

Machine Learning: El aprendizaje automático se puede definir como el área de la IA que se enfoca en el diseño, desarrollo y aplicación de algoritmos y modelos computacionales donde se mejora el rendimiento en función de la experiencia. Está destinada al análisis de patrones,

características e incidencias en grandes cantidades de datos. En pocas palabras, se refiere a la creación de programas computacionales que buscan mejorar en alguna métrica de evaluación utilizando datos. (Aracena et al., 2022)

Según Vega et al., (2020) los distintos métodos de aprendizaje automático se pueden clasificar en dos categorías según el enfoque de aprendizaje que utilicen:

- **Aprendizaje supervisado:** El propósito principal de estos algoritmos es aprender las reglas generales que mapean las entradas con las salidas enfocados en entrenamientos con datos relativamente simples y estructurados. Es decir que genera un modelo que utiliza un conjunto de entrenamiento para predecir una variable objetivo. (Aracena et al., 2022). Así pues, hay un proceso de retroalimentación constante en el modelo mediante la validación de datos de prueba correspondientes a la entrada y salida para marcar un precedente y así aprender a generar la predicción (Ilustración 3).

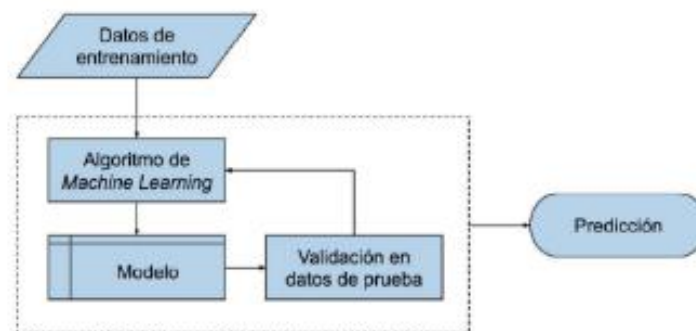


Ilustración 3 Esquema de aprendizaje supervisado.

Fuente: (Aracena et al., 2022)

- **Aprendizaje no supervisado:** A diferencia del aprendizaje supervisado, el algoritmos debe encontrar patrones y estructuras con similitud por sí mismo, este tipo de aprendizaje puede ser útil en un escenario dónde no se conozcan las etiquetas de los datos. Este algoritmo se basa en el entrenamiento de un modelo que utiliza datos sin etiquetas para realizar análisis descriptivo de los mismos como ser agrupaciones, asociaciones o detección de anomalías (Ilustración 4). (Aracena et al., 2022)

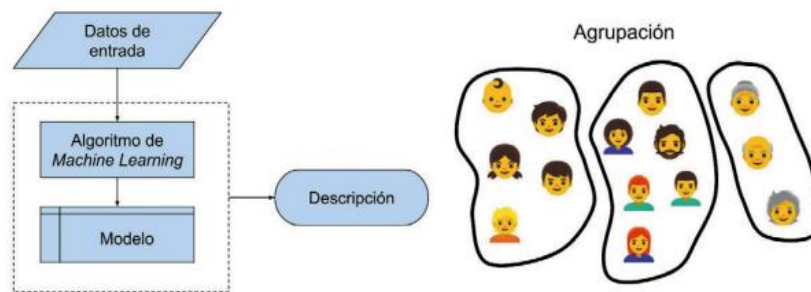


Ilustración 4 Esquema de aprendizaje no supervisado.

Fuente: (Aracena et al., 2022)

Así mismo Aracena et al. (2022) afirman que los algoritmos que más destacan para realizar aprendizaje automático han sido Random Forest, Support Vector Machines y Redes Neuronales. Adicionando el aprendizaje profundo o Deep Learning que se utiliza en aplicaciones más actuales y complejas.

Deep Learning: El DL, o aprendizaje profundo, se fundamenta en las redes neuronales artificiales convencionales, pero utiliza una gran cantidad de capas ocultas y neuronas, así como nuevos modelos y métodos de entrenamiento. Esto le brinda una mayor capacidad para el aprendizaje y adaptaciones receptivas para así lograr extraer características de datos de entrada de alta complejidad con mayor eficiencia. También se conocen como redes neuronales profundas. (Sarmiento-Ramos, 2020)

A diferencia de las técnicas de aprendizaje automático tradicionales, en el DL el modelo diseñado es capaz de realizar la tarea completa por sí solo. Es decir que se encarga de aprender y de extraer todos los datos junto con las características más relevantes de cada uno a partir de los datos de entrada.

Redes neuronales artificiales: Las redes neuronales artificiales están diseñadas con el propósito de replicar o asemejarse a las redes neuronales biológicas del cerebro humano. Las RNA son modelos computacionales que imitan la forma en que funcionan las células nerviosas en el cerebro humano. Estas utilizan algoritmos de aprendizaje que pueden hacer ajustes de forma independiente, o aprender, en cierto sentido, a medida que reciben nuevos datos. (Patiño et al., 2022)

Los términos inteligencia artificial, aprendizaje automático y aprendizaje profundo están estrechamente relacionados entre sí (Ilustración 5). Tanto el aprendizaje automático como el aprendizaje profundo se pueden incluir en una categoría más extensa como lo es la IA ya que ambos son métodos utilizados para lograr los objetivos de esta.

Así mismo, se involucran las redes neuronales artificiales dado que son sistemas implementados dentro del área del aprendizaje profundo donde las máquinas aprenden por sí solas.

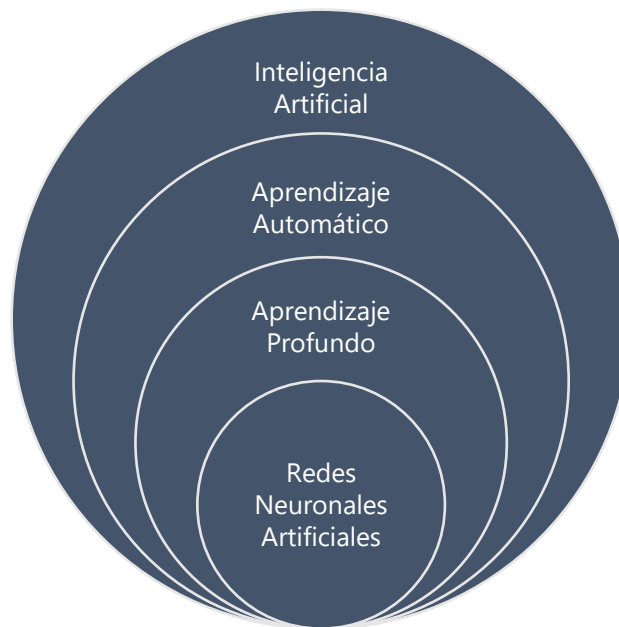


Ilustración 5 Relación de la inteligencia artificial con enfoques que involucran algún tipo de aprendizaje.

Fuente: (Choi et al., 2020)

Algoritmo: “Serie de instrucciones o rutina que permite entrenar un modelo, es decir, seleccionar los valores de los parámetros a partir de los datos del conjunto de entrenamiento.” (Aracena et al., 2022)

Árbol de decisión: El algoritmo de árbol de decisión es frecuentemente utilizado para realizar trabajos de regresión y clasificación mediante un análisis de datos cuantitativos. Esta es una técnica de aprendizaje supervisado en la que cualquier ruta que comience desde la raíz se describe mediante una secuencia de separación de datos hasta que se logra un resultado booleano en el nodo hoja. (Charbuty & Abdulazeez, 2021). Este trabaja mediante la

comparación entre una característica numérica con un valor de umbral en cada prueba para la toma de decisiones.

K-Nearest Neighbor Algorithm: El algoritmo de k-vecinos más cercanos es un método simple de aprendizaje automático. Consiste en asignar una clase a observaciones no etiquetadas a partir de la similitud con los ejemplos previamente etiquetados. (Zhang, 2016). La clasificación k-NN tiene dos etapas: la primera es la determinación de los vecinos más cercanos, y la segunda es la determinación de la clase usando esos vecinos. (Cunningham & Delany, 2022)

3.3 TEORÍAS DE SUSTENTO

En el siguiente apartado de sustento teórico se describieron algunas teorías, principios y fundamentos relacionadas con la inteligencia artificial y distintos métodos de aprendizaje automático. Tomando en cuenta también investigaciones previas y antecedentes en general para lograr situar correctamente el enfoque de la investigación.

3.3.1 BASES TEÓRICAS

Con la finalidad de brindar un enfoque orientado a explicar los métodos utilizados de aprendizaje automático, esta sección describe algunas leyes, principios y fundamentos necesarios para entender en su totalidad el proyecto.

3.3.1.1 FUNDAMENTOS DE LA INTELIGENCIA ARTIFICIAL

Los inicios de la inteligencia artificial se remontan a una inquietud filosófica de si una máquina puede imitar la naturaleza del pensamiento humano. En consecuencia, el estado actual de la inteligencia artificial no es más que el resultado de una larga y ardua búsqueda emprendida por el ser humano para crear seres artificiales capaces de realizar tareas de inteligencia, según su imagen y semejanza, o bien simulando el comportamiento de otras formas de vida. (Banda, 2014). La evolución de esta novedad tiene aplicabilidad en casi todos los ámbitos de la ciencia, incluyendo el diagnóstico clínico.

La inteligencia artificial pretende alcanzar 2 objetivos, uno de ellos relacionado con la parte tecnológica y el otro con la parte científica. Juntos, hacen referencia a que la IA busca utilizar ordenadores para ayudar a resolver cuestiones sobre los seres humanos y demás seres

vivos. (Boden, 2017). Teniendo en mente el cumplir esa meta, algunos de los logros de la IA hasta el momento se relacionan con agilizar procesos, realizar tareas complejas, analizar datos de forma rápida y efectiva reduciendo el error humano al máximo.

Con el paso de los años, la filosofía, en conjunto con la ingeniería, logró crear máquinas pensantes con capacidades resolutorias. Los primeros acercamientos a la IA surgen con los autómatas, que es hasta 1920 donde se les empieza a llamar robot, adoptando este nuevo termino. En julio de 1945, en la revista mensual The Atlantic, Vannevar Bush se publicó un artículo titulado "As We May Think", en el que se presentaba una visión de cómo las computadoras podrían asistir a los humanos en el futuro. (Serna, 2017)

Cómo toda disciplina de la ciencia, la IA tiene sus fundamentos en 4 principios que, según Serna (2017) son: capacidad de auto aprender, interacción con usuario, reacción en tiempo real y autonomía. De no contar con dichos principios, no sería considerada IA.

3.3.1.2 PRINCIPIOS DE LAS REDES NEURONALES ARTIFICIALES

Las redes neuronales artificiales están diseñadas para coincidir con las redes neuronales biológicas del cerebro humano. Las redes neuronales artificiales tienen sus bases y funcionan de una forma muy similar a las redes neuronales biológicas del cerebro de las personas. Su funcionamiento se da gracias a elementos que se comportan como una neurona biológica en sus funciones principales. (Serna, 2017)

Para comprender el funcionamiento de las RNA se debe comprender la función de las neuronas humanas, entre ellas se definen tres componentes principales: las dendritas, el cuerpo y el axón. El punto de conexión entre el axón de una célula y la dendrita de otra célula se llama sinapsis. La sinapsis es la conexión funcional que existe entre dos neuronas o entre una neurona y una célula. En términos computacionales, las dendritas son las receptoras artificiales de la red, que cargan el cuerpo celular con señales eléctricas. En el cuerpo de la célula se realiza la suma de las señales de entrada. Y por último el axón, una fibra larga que lleva la señal desde el cuerpo de la célula hasta otras neuronas, transmite las señales entre el conjunto de neuronas. (Serna, 2017)

3.3.1.2.1 TIPOS DE REDES NEURONALES

Existen diversos tipos de redes neuronales, cada uno con características específicas dirigidas a realizar acciones concretas. Se distinguen principalmente por la organización de sus capas y la función que efectúan.

3.3.1.2.1.1 MULTILAYER PERCEPTRÓN (MLP)

El tipo de RNA conocido como perceptrón multicapa está compuesto por múltiples capas de neuronas que se ubican entre la entrada y la salida, estas se denominan capas ocultas. La suma estas capas permite aumentar el potencial y las capacidades de la red para la solución de problemas que implican la clasificación y regresión con conjuntos de datos complejos. (Sarmiento-Ramos, 2020)

3.3.1.2.1.2 DEEP MULTILAYER PERCEPTRÓN (DMLP)

Las redes neuronales profundas que se basan en el perceptrón multicapa aplican la arquitectura de la red MLP, pero con un mayor número de neuronas y capas ocultas. Esto les permite resolver problemas donde los datos de entrada son de alta complejidad. (Sarmiento-Ramos, 2020). Este tipo de RNA suele estar compuesta por 2 o 3 capas ocultas con numerosas neuronas (Ilustración 6).

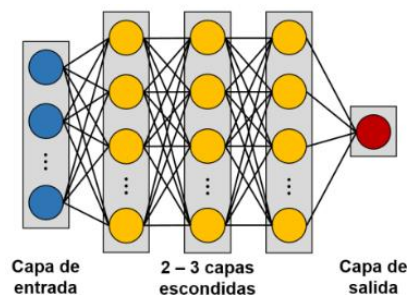


Ilustración 6 Organización de las capas del modelo DMLP.

Fuente: (Sarmiento-Ramos, 2020)

3.3.1.2.1.3 RED NEURONAL DE HOPFIELD

Este tipo de red neuronal funciona como una memoria asociativa en donde cada neurona está conectada a las demás (Ilustración 7). Durante el proceso de entrenamiento, se proporcionan todos los ejemplos de entrada, con la expectativa de que la red almacene y memorice estos patrones. Una vez finalizado el entrenamiento, la capacidad de memoria asociativa de la red se utiliza para recordar y reconocer los patrones almacenados, incluso en presencia de ruido. (Sarmiento-Ramos, 2020)

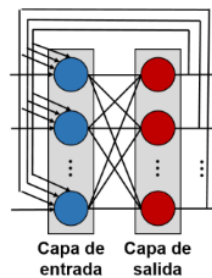


Ilustración 7 Organización de las capas del modelo Hopfield.

Fuente: (Sarmiento-Ramos, 2020)

3.3.1.2.1.4 SELF-ORGANIZING MAPS (SOMs)

Los mapas autoorganizados son un tipo de red cuyo funcionamiento se basa en tomar los datos de entrada y agruparlos de acuerdo a su similitud en las muestras o patrones. Están diseñadas para que las neuronas de salida con patrones similares estén ubicadas cerca unas de otras. (Sarmiento-Ramos, 2020). Para lograr esto, la organización en este tipo de red está diseñada con una única capa de entrada y una única capa de salida, a esta última también se le conoce como mapa de resultados o capa de Kohonen (Ilustración 8).

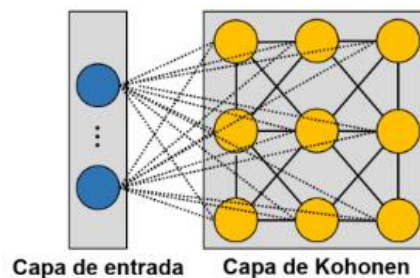


Ilustración 8 Organización de las capas del modelo SOMs.

Fuente: (Sarmiento-Ramos, 2020)

3.3.1.2.1.5 DEEP AUTO-ENCODER (DAE)

Las redes de tipo codificador automático profundo son un caso especial de DMLP. Estas tienen dos características principales: utilizan la misma cantidad de neuronas en la entrada y en salida, y emplean un menor número de neuronas en las capas ocultas con respecto a las utilizadas en la entrada y salida (Ilustración 9). (Sarmiento-Ramos, 2020)

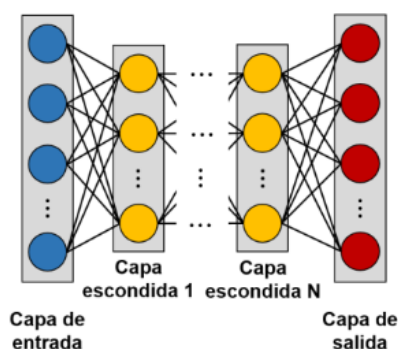


Ilustración 9 Organización de las capas del modelo DAE.

Fuente: (Sarmiento-Ramos, 2020)

3.3.1.2.1.6 RECURRENT NEURAL NETWORK (RNN)

Las redes neuronales recurrentes por su arquitectura, en esta las conexiones entre neuronas forman un gráfico dirigido a lo largo de una secuencia temporal (Ilustración 10). Esto les brinda una gran capacidad que les permite ser de gran utilidad para el análisis de datos en series de tiempo, además de demostrar un comportamiento dinámico y almacenar información sobre lo sucedido en periodos de tiempo anteriores gracias a que poseen una memoria interna. (Sarmiento-Ramos, 2020)

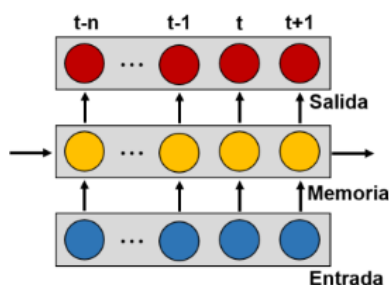


Ilustración 10 Organización de las capas del modelo RNN.

Fuente: (Sarmiento-Ramos, 2020)

3.3.1.2.1.7 RESTRICTED BOLTZMANN MACHINE (RBM)

La máquina restringida de Boltzmann es una RNA que, mediante un proceso aleatorio, es capaz de aprender una distribución de probabilidad a partir de un conjunto de datos de entrada. Su arquitectura se compone de una capa de neuronas visibles y una capa de neuronas ocultas que se conectan entre sí, impidiendo que existan conexiones entre las neuronas de un mismo tipo de capa (Ilustración 11). Según las características de este tipo de red, sus principales aplicaciones son el modelado de relaciones probabilísticas y el preentrenamiento de parámetros de DNNs de mayor complejidad. (Sarmiento-Ramos, 2020)

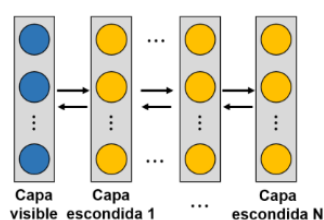


Ilustración 11 Organización de las capas del modelo RBM.

Fuente: (Sarmiento-Ramos, 2020)

3.3.1.2.1.8 CONVOLUTIONAL NEURAL NETWORK (CNN)

Las redes neuronales convolucionales están inspiradas en la corteza visual, por lo que principalmente, son utilizadas en tareas de clasificación o procesamiento de imágenes. Está formada por mapas del campo receptivo local, lo que les permite responder a los estímulos de una región específica del campo visual y disminuyen en tamaño a medida que se avanza en la corteza visual (Ilustración 12). Además, estos campos receptivos se superponen para lograr cubrir todo el campo visual. (Sarmiento-Ramos, 2020)

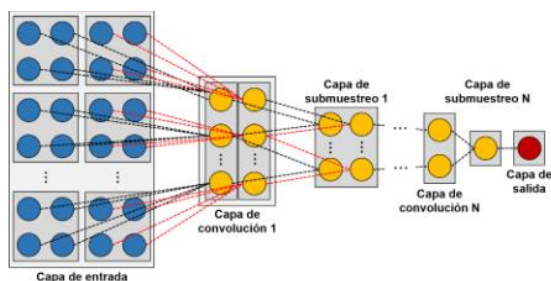


Ilustración 12 Organización de las capas del modelo CNN.

Fuente: (Sarmiento-Ramos, 2020)

3.3.1.2.2 NEURONA ARTIFICIAL

Se considera el centro de la construcción de una red neuronal artificial. La estructura de una neurona artificial es muy similar a una neurona biológica. Está formada por tres partes principales: peso y sesgo, como dendritas, denotados por w y b respectivamente, y la salida como un axón denotado por y . Además, se tiene una función de activación como un cuerpo celular (núcleo) denotado por $g(\bullet)$ (Ilustración 13). (Patiño et al., 2022)

En las neuronas artificiales, la entrada y el peso se representan como un vector, mientras que el sesgo se representa como un escalar. La neurona artificial procesa las señales de entrada realizando un producto punto entre el vector de entrada y el vector de peso, agrega el sesgo, luego aplica una función de activación y finalmente propaga el resultado a otras neuronas. (Patiño et al., 2022)

Según Ding et al. (2018), La neurona artificial recibe señales de las entradas (x_1, x_2, \dots, x_n) ponderadas por sus respectivos pesos sinápticos (w_1, w_2, \dots, w_n). Además, a estos se les añade un sesgo o bias (Ilustración 13). Esa sería la entrada neta o entrada total que recibiría la neurona artificial, donde se tiene una unidad de adición Σ que devuelve una señal z . Esta representa la suma de pesos lineales de las entrada con el sesgo.

Finalmente, a esa señal se le aplica una función matemática conocida como función de activación, representada por $g(\bullet)$, cuyo trabajo es estimular el estado de respuesta de la neurona para obtener la salida y .

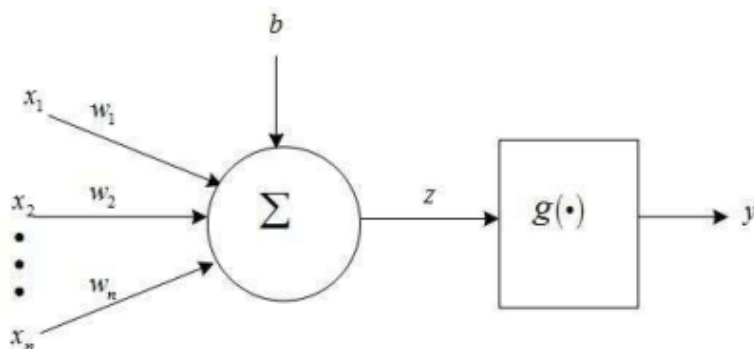


Ilustración 13 Diagrama de neurona artificial.

Fuente: (Ding et al., 2018)

3.3.1.2.3 CAPAS DE LAS NEURONAS

Las neuronas se componen por diversas capas, de manera que todas se encuentran interconectadas entre sí con el fin de compartir información. Según Serna (2017), cada neurona de la red es una unidad de procesamiento de información que recibe información a través de las conexiones con las neuronas de la capa anterior. Así mismo, expone la importancia de 3 tipos de capas:

- Capa de entrada: Esta es la encargada de recibir información proveniente del exterior, imitando la tarea de las dendritas en las redes biológicas.
- Capas ocultas: Se encargan de realizar el trabajo de la red, comparándolas con el sistema biológico, este trabajo correspondería al soma.
- Capa de salida: Brinda el resultado del trabajo realizado por la red al exterior y se encarga de enviar la información hacia otras neuronas, esta actividad sería realizada por el axón en el sistema biológico.

3.3.1.2.4 PESOS SINÁPTICOS

A las entradas provenientes de otras neuronas se les asigna un peso sináptico. El proceso es realizado con el propósito de imitar la sinapsis neuronal producida para la transmisión de información en el cuerpo humano. Se puede decir que el peso sináptico es entonces un valor numérico que puede cambiar durante la fase de entrenamiento. Este peso hace que la red neural tenga una utilidad y es allí donde se almacena la información. (Serna, 2017)

3.3.1.2.5 REGLA DE PROPAGACIÓN

En un modelo neuronal, se debe disponer de una regla de propagación para combinar las salidas de cada neurona con las ponderaciones establecidas por el patrón de conexión, con eso se especifica la valoración de las entradas que recibe cada neurona. Normalmente puede realizarse una suma de las entradas, teniendo en cuenta el peso sináptico asociado a cada entrada. Aunque otras operaciones también son posibles. (Serna, 2017)

Según Zajmi et al., (2018), la retropropagación es de las técnicas más utilizadas al desarrollar un modelo de red neuronal, en esta se tiene en cuenta una serie de elementos para incidir en su convergencia. El funcionamiento de esta regla de propagación básicamente se

realiza un cálculo de error que mide o compara las diferencias entre la salida deseada y la salida producida. Luego, el error es propagado hacia atrás entre las capas de la red, desde la salida a la oculta y a la entrada, modificando los pesos para reducir el error durante esta propagación.

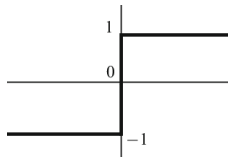
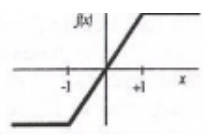
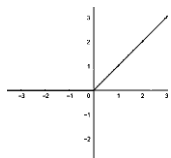
3.3.1.2.6 FUNCIÓN DE ACTIVACIÓN

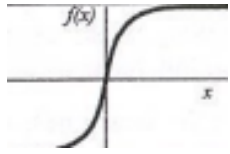
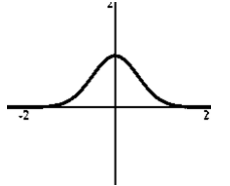
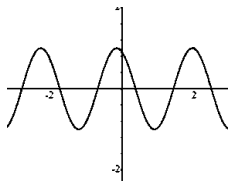
Con el valor obtenido con la regla de propagación, esta se filtra con una función conocida como función de activación. A través de esta función se establece la salida de la neurona. Las funciones de activación se seleccionan según el objetivo de entrenamiento de la red neuronal. Las funciones de activación más utilizadas son la sigmoidea, unidad lineal rectificadora (ReLU), escalón, lineal a tramos, gaussiana y sinusoidal (Tabla 3). (Serna, 2017)

Después de lo anterior expuesto, cabe resaltar que las funciones de activación evitan o permiten el paso de la información de una neurona a otra. Se puede decir que es una función limitadora a la salida de cada neurona que establece un umbral que se debe superar para que la información pueda ser transmitida.

La función sigmoidea es utilizada comúnmente en redes neuronales poco profundas. Esta función rara vez se emplea en las redes neuronales profundas, y se utiliza únicamente en el nivel de salida de la red neuronal debido a su distribución de valores, ya que está saturado. (Ding et al., 2018)

Tabla 3 Funciones de Activación más utilizadas.

	Función	Gráfica
Escalón	$y = \text{sgn}(x)$ $y = H(x)$	
Lineal a tramos	$y = \begin{cases} -1, & \text{si } x < -1 \\ x, & \text{si } -1 \leq x \leq +1 \\ +1, & \text{si } x > +1 \end{cases}$	
Unidad lineal rectificadora (ReLU)	$y = \begin{cases} 0, & \text{si } x < 0 \\ x, & \text{si } x \geq 0 \end{cases}$	

	Función	Gráfica
Sigmoidea	$y = \frac{1}{1+e^{-x}}$ $y = \operatorname{tgh}(x)$	
Gaussiana	$y = Ae^{-Bx^2}$	
Sinusoidal	$y = A \operatorname{sen}(wx + \theta)$	

Fuente: (Serna, 2017).

3.3.2 METODOLOGÍAS APLICADAS

En este apartado se exponen los métodos y procesos aplicados a investigaciones similares por otros expertos. Haciendo énfasis en el software utilizado, las bases de datos y las variables de investigación con el fin de marcar un punto de partida y esclarecer la ruta que se debe seguir en el transcurso de la investigación.

Rosales (2022) en su artículo estudia 13 posibles modelos para el análisis predictivo de la diabetes mediante el enfoque de aprendizaje automático, obteniendo como resultado que no todos los enfoques de IA son aptos para la obtención de un buen puntaje en las 7 métricas de diagnóstico en las que basó su investigación.

Pablo Rodríguez de León, et al. (2020) desarrollaron un software de clasificación y selección a través de dos mecanismos, uno basado en aprendizaje automático y el otro basado en la técnica de selección de testores. Como conclusión del estudio, afirman que las variables de estado civil son las que mayor influencia tuvieron, además de corroborar que los datos de laboratorio clínico y los antecedentes familiares también son influyentes en la eficacia de la red neuronal.

Castrillón et al. (2017), aplicando una metodología de 4 pasos y analizando bases de datos del instituto nacional de diabetes, lograron identificar que las variables de mayor influencia en el diagnóstico obtenido a través de la red neuronal son desde número de embarazos en mujeres hasta la edad y la herencia. Con estas presentaron un porcentaje de acierto del 87% analizando que se obtuvo un mejor porcentaje cuando se agrega la variable de insulina en suero logrando un desempeño del 98.46%. Como resultado, se pudo concluir que el clasificador bayesiano es funcional para el diagnóstico de la diabetes.

Patiño et al. (2022) estudiaron las diferentes técnicas de aprendizaje, utilizando el lenguaje de programación Python a través de google colab. Finalmente, pudieron concluir que el porcentaje de precisión obtenido fue del 99.3% logrando proporcionar un predictor por medio de redes neuronales artificiales analizando variables de análisis de laboratorio más datos demográficos y de estado civil.

La metodología utilizada está basada en técnicas de aprendizaje automático o Machine Learning al igual que una técnica de aprendizaje profundo o Deep Learning por medio de una red neuronal artificial, los pasos a seguirse se basaron en:

- Tratamiento de los Datos
- Creación del Modelo
- Fase de entrenamiento
- Fase de Prueba
- Evaluación del modelo con sus métricas
- Aplicación del modelo

(Patiño et al., 2022)

En su investigación se analizaron ocho características de pacientes objeto de análisis, estas fueron: número de embarazos, glucosa en plasma, presión sanguínea diastólica, espesor del pliegue cutáneo del tripces, insulina en suero, índice de masa corporal, función de pedigrí de la diabetes y la edad. Para el entrenamiento de la red tomó dos bases de datos, una de ellas corresponde a personas que padecen la enfermedad y la otra a personas que no la padecen.

Finalmente, obtuvo un porcentaje de precisión del 96.92% excluyendo el número de embarazos y la glucosa en plasma de las variables de entrada a ingresar al sistema.

Pineda (2022) en su investigación con fines de demostrar los avances tecnológicos orientados a la salud, habla sobre las aplicaciones de los modelos predictivos en las distintas áreas de la salud. En conclusión, el análisis menciona que en 2021 fue diseñada una librería del lenguaje de programación Python (el más utilizado en ciencia de datos y aprendizaje de máquina) para modelos predictivos en salud: PyHealth.

Esta consiste en módulos especiales para el preprocesamiento de datos, modelamiento predictivo (con 30 modelos disponibles) y de evaluación, de tal manera de diagramar modelos complejos de datos médicos de una forma simple y con pocas líneas de código. (Pineda, 2022)

3.4 MARCO LEGAL

El marco legal que encierra el uso de redes neuronales varía dependiendo de la región, estado o país así mismo como de las aplicaciones que se le den a las mismas. En general, existen leyes y regulaciones destinadas a la protección de la privacidad, la seguridad y los derechos de los datos personales, como también existen medidas para garantizar la calidad y la entereza de la información.

Uno de los aspectos destacables de la aplicación del marco de leyes a la inteligencia artificial es la Ley Orgánica de Protección de Datos y Garantía de Derechos Digitales (LOPD-GDD) y la protección de datos mediante el Reglamento General de Protección de Datos (RGPD).

Para el desarrollo del proyecto se tomaron en consideración el cumplimiento de las directrices establecidas anteriormente, además de otras normativas internacionales (Tabla 4) y nacionales (Tabla 5) que tienen gran importancia en el contexto confidencial del manejo de datos.

Tabla 4 Normativa Internacional.

Documento Legal	Entidad Reguladora	Directriz
Ley Orgánica 3/2018 (Protección de datos personales y garantía de derechos digitales)	Parlamento Europeo	Establece el supuesto de iuris tantum para la aceptabilidad del tratamiento de datos personales. Esto incluye la comunicación previa, o cesión de datos.
Real Decreto 69/2015	Registro de Actividad de Atención Sanitaria Especializada	Adoptar las medidas que garanticen la confidencialidad, seguridad e integridad de los datos contenidos en el registro.
ANSI/CTA-2090	Asociación de Tecnología del Consumidor (CTA)	Para que los usuarios confíen en una solución de IA, un pilar esencial es la seguridad de los datos. El desarrollador del modelo deberá cumplir con todas las normas de jurisdicción pertinentes relacionadas con la seguridad de los datos. Que estos sean confiables y seguros.
Carta de los Derechos Fundamentales de la Unión Europea Artículo 8.	Unión Europea	Toda persona tiene derecho a la protección de los datos de carácter personal que la conciernan.

Fuente: (OEA, 2022) (Ferrer, 2020) (ANSI/CTA Standard the Use of Artificial Intelligence in Health Care: Trustworthiness, 2021) (Olivares, 2018)

Localmente hablando, el marco legal de la privacidad se encuentra principalmente establecido en la Constitución de la República y la Ley de Protección de Datos Personales.

Tabla 5 Normativa Nacional.

Documento Legal	Entidad Reguladora	Directriz
Artículo 76.	Constitución de la República de Honduras.	Se garantiza el derecho al honor, a la intimidad personal, familiar y a la propia imagen.
Artículo 100.	Constitución de la República de Honduras.	Toda persona tiene derecho a la inviolabilidad y al secreto de las comunicaciones, en especial de las postales, telegráficas y telefónicas, salvo resolución judicial.
Artículo 182	Constitución de la República de Honduras.	<p>El Estado reconoce la garantía de Hábeas Corpus o Exhibición Personal, y de Hábeas Data. En consecuencia, en el Hábeas Corpus o Exhibición Personal toda persona agraviada o cualquier otra en nombre de ésta tiene derecho a promoverla; y en el de Hábeas Data únicamente puede promoverla la persona cuyos datos personales o familiares consten en los archivos, registros públicos o privados de la siguiente manera:</p> <ul style="list-style-type: none"> - EL HABEAS DATA. Toda persona tiene el derecho de acceder a la información sobre sí misma o sus bienes en forma expedita y no onerosa, ya esté contenida en bases de datos, registros Públicos o Privados y en el caso de que fuere necesario, actualizarla, rectificarla y o suprimirla. No podrá afectarse el secreto de las fuentes de información periodística. - Las acciones de Hábeas Corpus y Hábeas Data se deben ejercer sin necesidad de poder ni de formalidad alguna, verbalmente o por escrito, utilizando cualquier medio de comunicación, en horas o días hábiles e inhábiles y libres de costas.
Ley de protección de datos personales.	Instituto de Acceso a la Información Pública de Honduras.	Cada hondureño puede tener acceso a su información personal en posesión de cualesquiera terceros, ejerciendo un poder de elección sobre cuándo, cómo y quién va a tratar su información personal.

Fuente: (Constitución de la República de Honduras, 1982)

3.5 MATRIZ DEL ESTADO DEL ARTE

Todas las referencias utilizadas previamente se ven expuestas en una matriz donde se muestra el título de la investigación y el año de la misma, los autores y su campo de especialización, la ciudad o país donde esta fue realizada, y las conclusiones o resultados obtenidos (Tabla 6).

Tabla 6 Matriz del Estado del Arte.

Autores de la investigación	Título de la investigación	Año	Ciudad y país	Campo de especialización de los autores	Resultados/Conclusiones
Alam, U., Asghar, O., Azmi, S., & Malik	Chapter 15— General aspects of diabetes mellitus	2014	Amsterdam, Holanda	Endocrinología y Diabetes	N/A
Aracena, C., Villena, F., Arias, F., & Dunstan, J.	Aplicaciones de aprendizaje automático en salud	2022	Santiago, Chile	Ciencias Físicas y Matemáticas, Modelamiento Matemático	En un enfoque ético sobre el uso de la IA en salud se deben considerar 6 aspectos claves: <ul style="list-style-type: none"> • Proteger la autonomía humana • Promover el bienestar humano, la seguridad y el interés público • Garantizar la transparencia e inteligibilidad • Responsabilidad • Inclusividad y equidad • Sustentabilidad
Banda, H.	Inteligencia Artificial: Principios y Aplicaciones	2014	Quito, Ecuador	Informática y Ciencias Computacionales	N/A
Boden, M. A.	Inteligencia Artificial	2017	Madrid, España	Docente Investigador Ciencias cognitivas e IA	N/A
Bougherara, L., Hanssens, S., Subtil, D., Vambergue, A., &	Diabetes gestacional	2018	Francia	Endocrinología y Maternidad	Desde la estandarización de criterios para la detección precoz de diabetes gestacional, esta se ha convertido en una de las

Autores de la investigación	Título de la investigación	Año	Ciudad y país	Campo de especialización de los autores	Resultados/Conclusiones
Deruelle, P.					patologías más comunes del embarazo, sin el aporte de estudios prospectivos que demuestren beneficios para la salud de las mujeres y neonatos.
Campos, N., & Palomino, G.	FINDRISC, utility in the screening of diabetes, personalization, and associations	2018	Lima, Perú	Medicina	El 91% de las investigaciones recomienda el uso del test FINDRISC para la prevención primaria o como screening. El otro 9% reconoce su capacidad predictiva, resalta su potencial costo-beneficio, y sugiere estudios más rigurosos.
Carvalho, A. R. S., Guimarães, A., Werberich, G. M., de Castro, S. N., Pinto, J. S. F., Schmitt, W. R., França, M., Bozza, F. A., Guimarães, B. L. da S., Zin, W. A., & Rodrigues, R. S.	COVID-19 Chest Computed Tomography to Stratify Severity and Disease Extension by Artificial Neural Network Computer-Aided Diagnosis.	2020	Rio de Janeiro, Brasil	Medicina, Radiología, Ciencias Biomédicas, Ingeniería en Biomédica	El diagnóstico asistido por computadora basado en redes neuronales propuesto ayudó a detectar y cuantificar el grado de afectación pulmonar. Gracias a esto se logró fenotipar a los pacientes con neumonía por COVID-19 con un porcentaje del 81% de concordancia entre el diagnóstico del médico y el brindado por la red neuronal artificial.
Castrillón, O. D., Sarache, W., & Castaño, E.	Sistema Bayesiano para la Predicción de la Diabetes	2017	Colombia	Ingeniería Industrial, Medicina	Se propone un clasificador bayesiano para la identificación de personas con diabetes como una herramienta útil para la detección precoz de la enfermedad sin necesidad de pruebas de laboratorio. En particular, la precisión del sistema puede variar del 87,6 % al 96,9 %, según la cantidad de características de los pacientes analizadas.
Crespo, C., Brosa, M., Soria-Juan, A., Lopez-Alba, A., López-	Costes directos de la diabetes mellitus y de sus	2013	España	Estadística, Investigación Biomédica en red	Los costes relacionados con la DM suponen una proporción elevada del gasto sanitario total en España. Es necesario implementar

Autores de la investigación	Título de la investigación	Año	Ciudad y país	Campo de especialización de los autores	Resultados/Conclusiones
Martínez, N., & Soria, B.	complicaciones en España				estrategias y medidas para mejorar la eficiencia en el control y el tratamiento de la DM con el fin de reducir las complicaciones y los enormes costes humanos y económicos que conlleva. La mayor parte de los costos directos totales es debido a los costes farmacológicos (38%), seguido de los costes hospitalarios (33%).
Curioso, W. H., Brunette, M. J.	Inteligencia artificial e innovación para optimizar el proceso de diagnóstico de la tuberculosis.	2020	Perú	Informática Biomédica, Ingeniería Industrial y de Sistemas	Herramientas basadas en inteligencia artificial pueden llegar a optimizar procesos diagnósticos de la tuberculosis y otros problemas relacionados con la salud.
Dagogo-Jack, S.	Diabetes Mellitus in Developing Countries and Underserved Communities.	2017	Memfis, Tennessee, Estados Unidos	Docente de Medicina y Jefe de la División de Endocrinología, Diabetes y Metabolismo	Se estima que la prevalencia mundial de DM aumentará de 422 millones en 2014 a 592 millones en 2035. Desafortunadamente, se prevé que los países de economía baja y media experimentarán el mayor crecimiento, pero incluso en los países desarrollados, los subgrupos demográficos vulnerables muestran diferencias en la prevalencia, calidad de la atención y resultado del tratamiento.
Delgado, K., Ledesma, S., Rostro, H., Delgado, K., Ledesma, S., & Rostro, H.	Análisis de electroencefalograma usando redes neuronales artificiales.	2019	Salamanca, Guanajuato, México	Ingeniería Electrónica, División de ingenierías	Se desarrolló una RNA del tipo perceptrón multicapa para la clasificación de señales EEG de pacientes sanos, pacientes que controlaron las crisis y pacientes enfermos de epilepsia. Se realizó el entrenamiento de

Autores de la investigación	Título de la investigación	Año	Ciudad y país	Campo de especialización de los autores	Resultados/Conclusiones
ElSayed NA, Aleppo G, Aroda VR, et al	Classification and Diagnosis of Diabetes: Standards of Care in Diabetes—2023	2022	Estados Unidos	Medicina	la red aplicando el método LOOCV, siendo este el que proporciona los resultados más sobresalientes, productivos y confiables. N/A
Enderica, P. F. V., Mendoza, Y. O. G., Apolo, K. E. M., & Flores, J. J. O.	Diabetes Mellitus Tipo 2: Incidencias, Complicaciones y Tratamientos Actuales.	2019	Guayaquil, Ecuador	Medicina, Gestión de Servicios de Salud	La DM tipo 2 es considerada una de las enfermedades crónicas que mayor impacto tiene en la calidad de vida de la población mundial y es un verdadero problema de salud. La magnitud del problema, en cuanto a los gastos directos e indirectos, obliga a ser muy cuidadosos a la hora de elegir cuál es el mejor tratamiento para los pacientes. N/A
Ferrer, S. M.	La actuación de la ley de protección de datos española y el reglamento europeo: un estudio sobre su aplicación en el derecho a la información y el consentimiento informado, datos clínicos y documentos hospitalarios de los médicos y	2020	Brasil	N/A	N/A

Autores de la investigación	Título de la investigación	Año	Ciudad y país	Campo de especialización de los autores	Resultados/Conclusiones
	pacientes en España.				
Gellman, M. D.	Encyclopedia of Behavioral Medicine	2020	Miami, Florida, Estados Unidos	Fisiología	N/A
Herman, W. H.	The Global Burden of Diabetes: An Overview.	2017	Michigan, Estados Unidos	Medicina interna y epidemiología	La carga financiera de la diabetes es enorme. En 2015, la FID estimó que la DM representaba entre el 5 y el 20 % del gasto sanitario total en la mayoría de los países. El gasto global para tratar la DM y sus complicaciones se estimó en \$673 mil millones en 2015 y se espera que aumente a \$802 mil millones para 2040.
Jonkman, L. J., Ndungu, M., Connor, S. E., et al	A qualitative assessment of diabetes care access and disease management in Central America	2020	Pittsburgh, Estados Unidos	Medicina, Farmacia y Terapéutica	La mayoría de las personas en Honduras no tienen acceso a un seguro de salud y más de 1,5 millones no reciben atención médica de forma regular. Este estudio proporciona información clave sobre los desafíos y oportunidades para el control de la DM en las zonas rurales de Honduras.
Kaul, K., Tarr, J. M., Ahmad, S. I., Kohner, E. M., & Chibber, R.	Introduction to Diabetes Mellitus.	2013	Reino Unido	Ciencias Biomédicas y Clínicas	N/A
Loncaric, F., Camara, O., Piella, G., & Bijmens, B.	La integración de la inteligencia artificial en el abordaje clínico del paciente: Enfoque en la imagen cardiaca.	2021	Barcelona, España	Investigación Biomédica, Tecnologías de la Información y las Comunicaciones	Después de pruebas de concepto y posibles beneficios de la aplicación de la inteligencia artificial en el área de la cardiología, la visión de la integración completa de la IA en el tratamiento habitual del paciente cardiaco y el proceso de la información cardiológica está pasando a ser una realidad palpable.

Autores de la investigación	Título de la investigación	Año	Ciudad y país	Campo de especialización de los autores	Resultados/Conclusiones
Martínez Bernabé, E., Paluzie-Ávila, G., Terre Ohme, S., Ruiz Poza, D., Parada Aradilla, M. A., González Martínez, J., Albertí Valmaña, R., & Castellvi Gordo, M.	Sistemas de soporte a la toma de decisiones clínicas en insuficiencia renal.	2014	Girona, España	Departamento de Informática, Servicio de farmacia y Documentación Médica	Los sistemas de soporte a la toma de decisiones clínicas y las órdenes médicas electrónicas obtuvieron una aceptación inmediata de la intervención automatizada (éxito) en el 65,69% de los casos.
Ocaña-Fernández, Y., Valenzuela-Fernández, L. A., & Garro-Aburto, L. L.	Inteligencia artificial y sus implicaciones en la educación superior.	2019	Lima, Perú	Docente Investigador	N/A
Pariante Rodrigo, E., Deib-Morgan, K., García de Diego, O., García-Velasco, P., Sgaramella, G. A., & García González, I.	Exactitud y concordancia entre glucómetros: Un estudio en condiciones habituales de práctica clínica	2017	España	Medicina Familiar, Salud Pública y Atención Primaria	Los glucómetros tienden a ser muy precisos, y en la práctica, la glucemia capilar y la glucemia plasmática se utilizan indistintamente. Sin embargo, diversas variables pueden afectar la eficacia de estos dispositivos. Los 3 glucómetros estudiados en la investigación no cumplieron con el estándar actual de la Asociación Americana de Diabetes.
Patiño, D., Iñiguez, F., & Nivelá, M.	Modelos de Machine Learning basados en Aprendizaje Supervisado para la Detección de Diabetes Mellitus en la Ciudad de Guayaquil.	2022	Guayaquil, Ecuador	Ciencias Matemáticas y Física	El uso un predictor implementado con redes neuronales será de gran ayuda en el sector salud debido a que puede diagnosticar la presencia o ausencia de la DM en una persona en cualquier instante. Es capaz de revisar un gran número de exámenes efectivamente y, así mismo, las instituciones sanitarias podrían brindar un mejor servicio ajustándose a las políticas

Autores de la investigación	Título de la investigación	Año	Ciudad y país	Campo de especialización de los autores	Resultados/Conclusiones
Pineda, J. M.	Modelos predictivos en salud basados en aprendizaje de maquina (machine learning).	2022	Santiago, Chile	Ciencia de datos	públicas para agilizar el proceso de diagnóstico sin afectar el presupuesto del centro. Actualmente, en combinación con técnicas modernas de aprendizaje automático, diversas fuentes de datos clínicos nos permiten generar rápidamente o en tiempo real modelos de predicciones para miles de tareas o problemas clínicos similares. En salud, el tiempo del diagnóstico es un factor crucial en el pronóstico del paciente.
Poveda, K. A. F., García, K. J. Q., Subía, D. L. F., & Choez, C. A. C.	Utilidad de hemoglobina glicosilada en diabetes tipo 2.	2020	Guayaquil, Ecuador	Medicina Interna, Endocrinología	Actualmente, la HbA1c es la prueba de elección para el seguimiento crónico y control de la DM. En el control del diabético tipo 2, la medición de la HbA1c es una herramienta clave que brinda un gran apoyo para este proceso.
Richesson, R. L., Rusincovitch, S. A., Wixted, D., et al	A comparison of phenotype definitions for diabetes mellitus.	2013	Estados Unidos	Investigación en Informática Aplicada, Endocrinología, Metabolismo y Nutrición	Tres definiciones fenotípicas basadas principalmente en los códigos ICD-9-CM identificaron entre el 9% y el 11% de la población de pacientes. La definición para la Durham Diabetes Coalition incluye criterios adicionales e identifica el 13%. Las características demográficas para las siete definiciones fenotípicas fueron similares.
Ríos, B. R.	Diagnóstico del dengue utilizando redes neuronales artificiales.	2022	Tarapoto, Perú	Ingeniería de Sistemas e Informática	Utilizando una RNA-MLP, tipo cascade feed forward, con arquitectura de red 24 neuronas en la capa de entrada, 12 neuronas en la primera capa oculta, 4 neuronas en la segunda capa oculta y una neurona en la capa de salida, con función

Autores de la investigación	Título de la investigación	Año	Ciudad y país	Campo de especialización de los autores	Resultados/Conclusiones
Röder, P. V., Wu, B., Liu, Y., & Han, W.	Pancreatic regulation of glucose homeostasis.	2016	Singapore	Metabolismo en enfermedades humanas, Medicina metabólica	de activación tansigen las capas ocultas y de salida, algoritmo de entrenamiento trainlm (Levenberg Marquardt), se obtuvo un alto porcentaje de efectividad para el diagnóstico del dengue a partir de los síntomas de la enfermedad en la región San Martín, con un a 90.175 % El páncreas tiene un papel importante en el mantenimiento de niveles normales de glucosa en la sangre al producir y liberar insulina y glucagón. Las deficiencias en la secreción y/o la sensibilidad a la insulina pueden dar lugar a enfermedades metabólicas como la DM.
Sarmiento-Ramos, J. L.	Aplicaciones de las redes neuronales y el deep learning a la ingeniería biomédica.	2020	Colombia	Ingeniería Biomédica	En la gestión y administración de la salud pública, las principales aplicaciones de las redes neuronales se centran en el desarrollo de sistemas para el seguimiento y modelado de enfermedades en establecimientos de salud o largas poblaciones, el seguimiento y predicción de enfermedades asociadas a hábitos y estilos de vida usando información contenida en redes sociales, así como monitorear y pronosticar la calidad del aire.
Schnell, O., Barnard, K., Bergenstal, R., et al	Clinical Utility of SMBG: Recommendations on the Use and Reporting of SMBG in Clinical Research.	2015	Estados Unidos	Medicina Interna, Diabetes, Endocrinología y Metabolismo	N/A

Autores de la investigación	Título de la investigación	Año	Ciudad y país	Campo de especialización de los autores	Resultados/Conclusiones
Serna, E.	Principios y características de las redes neuronales artificiales.	2017	Medellín, Colombia	Investigación Científica, Ingeniería y Ciencia	N/A
Sprockel, J., Diaztagle, J., Alzate, W., & Gonzales, E.	Redes neuronales en el diagnóstico del infarto agudo de miocardio.	2014	Colombia	Medicina Interna, Ingeniería en Sistemas, Ingeniería Eléctrica	Es viable entrenar una red neuronal a partir de los parámetros de la escala de clasificación de probabilidad de Braunwald para el diagnóstico del IAM, con un desempeño tal que podría llegar a servir para el apoyo de la toma de decisiones de los médicos de urgencias.
Vega, M. Á., Mora, L. M. Q., & Badilla, M. V. C.	Inteligencia artificial y aprendizaje automático en medicina.	2020	Costa Rica	Medicina y Cirugía	En los últimos años con la creciente disponibilidad de información clínica almacenada electrónicamente el campo médico se ha convertido en un ambiente ideal para el desarrollo y aplicación de estas nuevas tecnologías. El aprendizaje automático posee el potencial de mejorar los sistemas de salud, mediante el análisis de millones de datos clínicos se logran crear modelos pronósticos, de tamizaje y diagnósticos.

Fuente: Elaboración propia.

IV. METODOLOGÍA

Este capítulo detalla la metodología aplicada para el desarrollo de la investigación. Donde se especifica el enfoque que la misma posee, las variables de investigación, las técnicas e instrumentos aplicados. Al mismo tiempo se describe la metodología de estudio y metodología de validación aplicadas.

4.1 ENFOQUE

La investigación se desarrolló con un enfoque cuantitativo, diseño experimental de tipo preexperimental con una secuencia de estudio transversal y alcance descriptivo (Tabla 7).

Se define que tiene un enfoque cuantitativo debido a que al finalizar se realizó un análisis de datos estadístico para la validación del funcionamiento de la red. Por ese mismo hecho de recolectar datos para medir las variables, se define que posee un alcance descriptivo.

Así mismo, se estableció como un estudio transversal ya que se realiza la investigación durante un periodo de tiempo determinado y breve. Y, finalmente, el diseño experimental se divide en 3 tipos: preexperimental, cuasi experimental y experimental puro. Para este caso, se utilizó un diseño preexperimental debido a que a estos se les considera como ensayos de otros experimentos con mayor control. (Sampieri, 2014)

Tabla 7 Metodología de la Investigación.

Metodología de la Investigación	
Enfoque	Cuantitativo
Estudio	Transversal
Diseño	Preexperimental
Alcance	Descriptivo

Fuente: Elaboración propia.

4.2 VARIABLES DE INVESTIGACIÓN

Después de determinar el enfoque de la investigación se plantearon las variables de investigación, clasificadas en variable dependiente y variables independientes (Ilustración 14). Mismas que se consideran de mayor influencia para obtener los resultados esperados al finalizar la investigación.

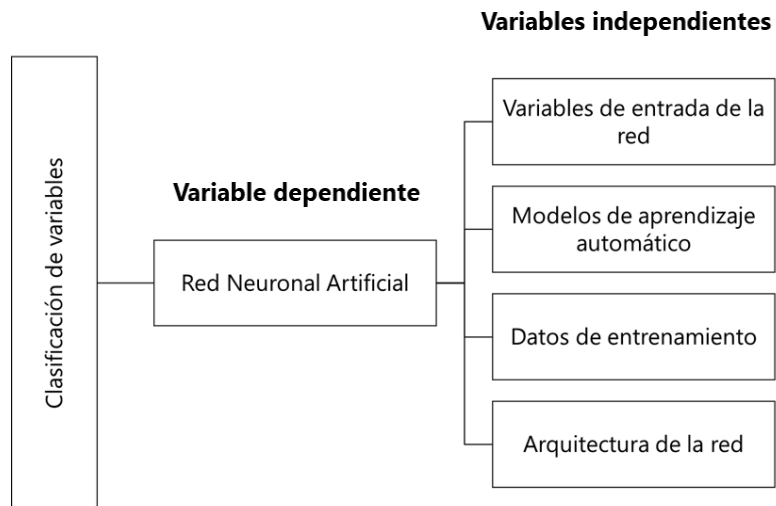


Ilustración 14 Identificación de variable dependiente y variables independientes.

Fuente: Elaboración propia.

4.2.1 VARIABLE DEPENDIENTE

Según Sampieri (2014), la variable dependiente es aquella que no se manipula, sino que se mide para identificar el efecto que genera la manipulación de las variables independientes sobre ella. En este caso, la variable dependiente es el desarrollo de una red neuronal artificial para el diagnóstico preciso y predicción de DM.

4.2.2 VARIABLES INDEPENDIENTES

La variable dependiente se ve afectada por un conjunto de variables independientes:

1. Variables de entrada de la red: La cantidad de variables o neuronas de entrada utilizadas en la red tienen un gran impacto en el porcentaje de precisión de la misma. El número de variables va a depender directamente de la cantidad de características con la que se

quiera entrenar a la red, en este caso, resultados de exámenes de laboratorio y preclínica.

2. Modelos de aprendizaje automático: Existen modelos de aprendizaje automático más precisos que otros según la tarea que se desea realizar. En el caso de trabajos relacionados a imágenes se suelen utilizar redes neuronales convolucionales, mientras que para clasificación mediante comparaciones de valores numéricos un método usado frecuentemente es el modelo de árbol de decisión. Por tal motivo, diversos métodos de aprendizaje automático pueden ser más funcionales que otros para comparar los resultados generados por la red neuronal y evaluar el rendimiento de la misma.
3. Datos de entrenamiento: La base de datos de una red neuronal es sumamente necesaria para el entrenamiento de esta. Las redes neuronales artificiales deben ser alimentadas con una amplia base de datos de entrenamiento para tomar como punto de referencia para el aprendizaje. Mientras mayor sea la base de datos con la que se alimente a la red, mayor será la cantidad de información que esta aprende y, por lo tanto, menor será el porcentaje de error generado.
4. Arquitectura de la red: La arquitectura de las redes neuronales artificiales se basa en cómo está organizada. Existen diversos tipos de arquitecturas dependiendo de la cantidad de capas, nodos o neuronas y sus características. (Sandoval-Ruiz, 2019).

4.3 TÉCNICAS E INSTRUMENTOS APLICADOS

Se detallan en esta sección los diversos instrumentos y técnicas que fueron utilizados en el proceso de elaboración de la red neuronal. Además del entorno de desarrollo integrado para la programación de la red, se acudió a un software para procesos estadísticos de validación de la red y plataformas para la recopilación de datos (Tabla 8).

Tabla 8 Instrumentos utilizados.

Tipo	Instrumento
Software	Visual Studio Code Minitab
Recolección de datos	Microsoft Excel

Fuente: Elaboración propia.

4.3.1 VISUAL STUDIO CODE

Visual Studio Code es un ligero pero potente editor de código fuente que cuenta con soporte integrado para JavaScript, TypeScript y Node.js y un amplio ecosistema de extensiones para otros lenguajes de programación como ser C++, C#, Java y Python. (*Visual Studio*, s. f.)

Este es conocido por su sintaxis clara y legible, lo que lo hace fácil de manejar. Además, el lenguaje de programación Python tiene una amplia variedad de librerías y módulos disponibles, lo que lo hace adecuado para una amplia gama de aplicaciones, desde la automatización hasta la ciencia de datos y el desarrollo de aplicaciones web e inteligencia artificial.

4.3.1.1 PANDAS

Pandas es una librería bastante útil para el análisis de datos en Python cuya función se basa en representaciones de estos conjuntos de datos. Esta permite la lectura y escritura de datos en diferentes formatos, selección, filtrado, unión y manipulación de datos, entre otras cosas. Suele ser utilizada en conjunto con la librería NumPy para la representación de datos.

4.3.1.2 NUMPY

La librería NumPy posee importantes funciones en cuanto al cálculos numéricos y arreglos matemáticos. Proporciona operaciones matemáticas integrales, generadores de números aleatorios, rutinas de álgebra lineal, transformadas de Fourier y más. (*NumPy*, s. f.). Debido al componente matemático involucrado en las redes neuronales artificiales, es una librería bastante útil cuando se trata del desarrollo y entrenamiento de las mismas.

4.3.1.3 SEABORN

Seaborn es una librería de visualización de datos de Python basada en matplotlib. Esta brinda una interfaz de alto nivel para la demostración de gráficos estadísticos atractivos e informativos. (Waskom, 2021). Permitiendo también la modificación y ajuste de ejes, colores, tamaños y demás para hacer el gráfico más entendible y agradable a la vista.

Es de gran utilidad para realizar mapas de calor y demás gráficos cuyo objetivo es mostrar la relación que existe entre cada una de las variables de entrada de la RNA con el resultado esperado.

4.3.1.4 MATPLOTLIB

Matplotlib simplifica la creación de visualizaciones estáticas, animadas e interactivas desarrolladas en el lenguaje de programación Python. (*Matplotlib — Visualización con Python*, s. f.). Este abarca desde la creación de gráficos simples como gráficos de dispersión y de barra, hasta aquellos más complejos como parcelas de matrices y campos o gráficos 3D.

En cuanto a su uso en análisis estadísticos, permite el desarrollo de diagramas de caja, histogramas, diagramas de eventos, entre otros.

4.3.1.5 KERAS

Keras es una biblioteca de aprendizaje profundo de código abierto escrita en Python. Fue desarrollada para facilitar la creación, entrenamiento y evaluación de modelos de aprendizaje profundo, especialmente redes neuronales, de una manera más rápida y eficiente. (Keras, 2020)

Keras se enfoca en la modularidad y la facilidad de uso, lo que la hace popular entre los investigadores y desarrolladores de aprendizaje profundo. Proporciona una API de alto nivel que permite la construcción de modelos de aprendizaje profundo con pocas líneas de código, lo que hace que sea fácil experimentar con diferentes arquitecturas de red y configuraciones de entrenamiento. (Keras, 2020)

4.3.1.6 TENSORFLOW

TensorFlow es una plataforma de software de código abierto de aprendizaje automático y de inteligencia artificial desarrollada por Google. Es una biblioteca de software diseñada para simplificar y agilizar el desarrollo de modelos de aprendizaje automático y redes neuronales para una amplia gama de aplicaciones, incluyendo el procesamiento del lenguaje natural, la visión artificial, la robótica, la detección de fraudes y muchas otras. (TensorFlow, 2020)

TensorFlow proporciona una serie de herramientas y API que permiten a los desarrolladores crear modelos de aprendizaje automático de manera eficiente y flexible, incluyendo la creación de redes neuronales profundas y la ejecución de entrenamientos de gran escala en sistemas distribuidos. También ofrece soporte para varios lenguajes de programación, incluyendo Python, Java, C++ y R, lo que facilita su integración en diferentes entornos de desarrollo. (TensorFlow, 2020)

4.3.1.7 Cv2

Cv2 es una biblioteca de procesamiento de imágenes y visión por computadora en Python. Es una de las bibliotecas más utilizadas en el campo de la visión por computadora y se utiliza ampliamente para tareas como el análisis de imágenes, el reconocimiento facial, la detección de objetos, la calibración de cámaras y la visión estereoscópica. (Visión artificial, OpenCV y Python, 2016)

La biblioteca cv2, también conocida como OpenCV (Open Source Computer Vision Library), es de código abierto y se puede usar en una amplia variedad de plataformas, como Windows, Linux, macOS y Android. Proporciona una serie de herramientas y funciones para trabajar con imágenes y videos, como la manipulación de imágenes, el filtrado y la segmentación, la detección de bordes y características, la correspondencia de puntos y la transformación de perspectiva. (Visión artificial, OpenCV y Python, 2016)

4.3.1.8 TKINTER

Tkinter es una biblioteca estándar de Python para crear interfaces gráficas de usuario (GUI, por sus siglas en inglés) de manera rápida y sencilla. Es una interfaz de Python para la biblioteca gráfica Tk, que fue desarrollada originalmente para el lenguaje de programación Tcl/Tk. (tkinter — Python interface to Tcl/Tk, 2023)

Tkinter se utiliza para crear ventanas, cuadros de diálogo, botones, menús, cuadros de texto, etiquetas y otros elementos de la interfaz de usuario. La biblioteca proporciona una serie de widgets y herramientas para diseñar y personalizar la apariencia de las interfaces de usuario. (tkinter — Python interface to Tcl/Tk, 2023)

4.3.1.9 SCIKIT-LEARN

Scikit-learn es una biblioteca de aprendizaje automático de código abierto para Python. Es una de las bibliotecas más utilizadas en el campo del aprendizaje automático y es muy popular entre los científicos de datos y los desarrolladores de aprendizaje automático. (Scikit-learn machine learning in Python — scikit-learn 1.2.2 documentation, s. f.)

Scikit-learn proporciona una amplia gama de herramientas y algoritmos para trabajar con datos, incluyendo clasificación, regresión, clustering, selección de características, reducción de dimensionalidad, preprocesamiento de datos y validación de modelos. La biblioteca también es altamente optimizada para grandes conjuntos de datos y es compatible con otras bibliotecas de Python como NumPy, SciPy y Pandas. (Scikit-learn machine learning in Python — scikit-learn 1.2.2 documentation, s. f.)

4.3.2 MINITAB

Minitab es un software estadístico bastante conocido, desarrollado en 1972 por tres profesores universitarios como herramienta de enseñanza para cursos de introducción a la estadística. (Alin, 2010). Para fines de esta investigación se utiliza como un análisis estadístico descriptivo donde se evalúen parámetros como ser la media, varianza, desviación estándar, entre otros con el fin de realizar conclusiones a partir de los datos obtenidos.

4.3.3 MICROSOFT EXCEL

Microsoft Excel es un software de hoja de cálculo desarrollado por Microsoft. Es una herramienta que permite la creación, organización y manipulación de datos numéricos y alfanuméricos en forma de tablas y gráficos. Crea hojas de cálculo con facilidad a partir de plantillas o desde cero y realiza cálculos con fórmulas modernas (Software de hojas de cálculo Microsoft Excel | Microsoft 365, s. f.)

Excel es ampliamente utilizado en entornos empresariales, educativos y personales debido a su capacidad para simplificar y automatizar proceso y su versatilidad en el manejo de datos. Es por eso por lo que fue utilizado con fines de almacenar los datos de las variables de entrada recopilados de los expedientes médicos de los pacientes.

4.4 MATERIALES

Los materiales requeridos para el desarrollo de la red neuronal artificial fueron la base de datos de entrenamiento y de prueba obtenidas a partir de los expedientes de pacientes del Hospital El Progreso para el entrenamiento y verificación del funcionamiento de la red y una computadora con los softwares necesarios para la programación de la misma.

4.5 POBLACIÓN Y MUESTRA

La población para esta investigación está determinada por el Hospital El Progreso, un establecimiento de salud público ubicado en la ciudad de El Progreso, Yoro, Honduras.

Se aplicó un tipo de muestreo no probabilístico, y se utilizó un cálculo de muestra con población desconocida para determinar qué tan amplia debe ser la base de datos a utilizar (Ecuación 1).

$$n = \frac{Z_{\alpha}^2 \times p \times q}{e^2}$$

Ecuación 1 Fórmula para cálculo de muestra de tipo no probabilístico con población desconocida.

Fuente: Elaboración propia.

Siendo:

n: Tamaño de la muestra.

Z α : Nivel de confianza (Tabla 9).

p: Proporción de la población que tiene el atributo deseado.

q: Proporción de la población que no tiene el atributo deseado (1-p).

e: Error de estimación máximo aceptado.

Tabla 9 Valores de Z.

Nivel de confianza deseado	Puntuación Z
90%	1.65
95%	1.96
98%	2.33
99%	2.58

Fuente: (Mendenhall et al., 2009)

Para el cálculo de muestra se tomó un nivel de confianza del 95%, por tal motivo, la Puntuación Z es de 1.96. En el caso de la proporción de población que tiene el atributo deseado se consideró el 62%, en consecuencia, para la población que no tiene el atributo deseado se tomó el 38%. Finalmente, se tomó en cuenta 7.5% para el error máximo aceptado y al aplicar todos estos parámetros en la fórmula se obtuvo una muestra total de 161 personas (Ecuación 2).

$$n = \frac{1.96^2 \times 0.62 \times 0.38}{0.075^2} = 161$$

Ecuación 2 Aplicación de fórmula para cálculo de muestra.

Fuente: Elaboración propia.

Se recopiló una base de datos total de 170 registros, lo cual cumple con lo establecido mediante el cálculo de muestra realizado. Estos se dividen en 65 pacientes del área de consulta externa sin un diagnóstico dado de DM, dentro de los cuales se encuentran también pacientes prediabéticos, y 105 pacientes de la clínica de diabetes del Hospital El Progreso diagnosticados con DM, en su mayoría DM2.

4.6 METODOLOGÍA DE ESTUDIO

Para el desarrollo de la investigación se utilizó como herramienta la metodología en V, misma que es útil para proyectos complejos que requieren una planeación exhaustiva y orientación sistemática que garantiza el éxito de los proyectos. Esta metodología se representa gráficamente en forma de V para describir la secuencia de las fases del proyecto de investigación (Ilustración 15).

4.6.1 FASE DE DISEÑO

La fase de verificación de la metodología en V abarcó de la realización de una investigación en cuanto a las variables asociadas a la DM, definición de las variables que se utilizarán como entrada de la RNA, definición del lenguaje de programación e IDE y finalmente, definición de la arquitectura de la red.

4.6.1.1 INVESTIGACIÓN DE VARIABLES ASOCIADAS

Con el objetivo de identificar las variables de interés para el estudio, se realizó una revisión bibliográfica de investigaciones similares para establecer las variables utilizadas. Así mismo se revisaron los expedientes médicos de los pacientes diabéticos para determinar la información relevante en ellos creando un panorama de la ruta a seguir. Finalmente, se charló del tema con expertos de la diabetes para recopilar toda la información necesaria para aclarar el camino.

4.6.1.2 DEFINICIÓN DE VARIABLES DE ENTRADA PARA LA RED NEURONAL

Luego del proceso de investigación por medio de la revisión bibliográfica, análisis de expedientes y charlas con expertos, se definieron las 10 variables que serán consideradas para el diagnóstico y predicción de DM por medio de la red neuronal (Tabla 10).

Según (American Diabetes Association, 2022) y (OPS/OMS-Secretaría de Salud de la República de Honduras, 2015), la prevalencia de la DM aumenta con la edad. Por tal motivo, se recomienda realizar pruebas a partir de los 35 años en caso de no presentar factores de riesgo como ser sobrepeso u obesidad, hipertensión arterial, niveles elevados de triglicéridos y

colesterol, antecedentes familiares, síndrome de ovario poliquístico en el caso de las mujeres o que hayan padecido diabetes gestacional.

Además, la hemoglobina glicosilada (HbA1c, o simplemente A1c) es una variable sumamente importante ya que por encima de 6.5% (48 mmol/mol) se considera paciente diabético y entre 5.7–6.4% (39–47 mmol/mol) paciente prediabético. Los valores normales están por debajo de los 5.7%. Esta representa el nivel promedio de glucosa en sangre en los últimos 3 meses.

Por otro lado, la glucosa en sangre, específicamente glucosa plasmática en ayunas es un valioso factor para el diagnóstico de la patología. Los prediabéticos tienden a obtener resultados entre 100–125 mg/dL (5.6–6.9 mmol/L), mientras que los diabéticos obtienen valores superiores a 126 mg/dL (7.0 mmol/L).

Sobre la base de las consideraciones anteriores, se consideraron de gran importancia para el diagnóstico de DM las siguientes variables: edad, peso, hemoglobina glicosilada, colesterol, triglicéridos, y presión arterial diastólica y sistólica.

En el caso del sexo, esta variable fue tomada en consideración para el estudio ya que en diversas investigaciones se ha demostrado que la DM se da más frecuentemente en las mujeres. (Agudelo-Botero & Dávila-Cervantes, 2015)

Tabla 10 Identificación de variables de entrada de la red neuronal.

Variable de entrada	Descripción	Tipo de dato
Edad	Edad en años	Numérico
Peso	Peso en libras	Numérico
Sexo	(1=Mujer, 2=Hombre)	Categorico
Glucosa en sangre	Valor en miligramos por decilitro (mg/dL)	Numérico
Hemoglobina glicosilada	Valor en porcentaje (%)	Numérico
Colesterol total	Valor en miligramos por decilitro (mg/dL)	Numérico
Presión arterial sistólica	Valor en milímetros de mercurio (mmHg)	Numérico
Presión arterial diastólica	Valor en milímetros de mercurio (mmHg)	Numérico
Triglicéridos	Valor en miligramos por decilitro (mg/dL)	Numérico
Creatinina	Valor en miligramos por decilitro (mg/dL)	Numérico
Target	Diagnóstico de DM (1=Verdadero, 0=Falso)	Target

Fuente: Elaboración propia.

4.6.1.3 SELECCIÓN DEL LENGUAJE DE PROGRAMACIÓN E IDE

Para la creación de la red, se concretó un proceso riguroso de búsqueda entre los diferentes softwares de programación, logrando concluir con ayuda de expertos que el lenguaje que más se presta por su popularidad en el campo de IA es Python. Entre las razones que pesan para la selección de este lenguaje se encuentran las diferentes librerías y bibliotecas que posee, es fácil de aprender y es flexible por lo que colabora en la creación de códigos de manera rápida, sencilla y eficiente.

4.6.1.4 DEFINICIÓN DE ARQUITECTURA DE SOFTWARE Y MÉTODOS DE ML

Se estableció la realización del código bajo una secuencia lógica específica apoyada con la arquitectura de las librerías TensorFlow y Keras (Ilustración 23). El modelo de red neuronal se diseñó bajo la estructura de 3 capas. La primera con 32 neuronas, la segunda es una capa dropout, por último, una capa densa de salida con una única neurona. A la primera se le aplica la función de activación reLU (Anexo 1) y finalmente, se tiene una capa de salida con función de activación Sigmoidea donde se definirá el diagnóstico y porcentaje predictivo de padecimiento de DM.

En cuanto a los métodos de aprendizaje automático considerados se tomaron el modelo de árbol de decisiones y el algoritmo de k-nearest neighbor. Se evaluó el rendimiento de cada uno de estos para determinar cuál sería el ideal para comparar los resultados obtenidos con los de la red neuronal y así poder realizar la validación.

4.6.2 FASE DE VALIDACIÓN

La fase de validación inició una vez realizada la programación de la red neuronal artificial. Se basó en el entrenamiento de los distintos métodos de inteligencia artificial utilizados, la validación de coherencia entre los resultados obtenidos, el análisis de los valores predictivos de la RNA y, finalmente, el desarrollo de la interfaz gráfica.

4.6.2.1 ENTRENAMIENTO DE LOS MÉTODOS DE INTELIGENCIA ARTIFICIAL

Para el entrenamiento de los métodos de inteligencia artificial utilizados, se desarrollaron dos algoritmos más, además de la red neuronal, en lenguaje de programación Python con el fin

de obtener un punto de comparación entre ellos. Para la red neuronal se utilizó el algoritmo de optimización Adam el cual funciona mediante una estimación adaptativa de momentos y es el optimizador ideal para experimentos rápidos.

Para el entrenamiento del árbol de decisión y k-nearest neighbor se utilizaron los mismos datos que en el entrenamiento de la red. Todo esto con el objetivo de que la entrada de cada uno de los métodos sea lo más parecido posible para así lograr establecer una comparación en la salida de los mismos.

4.6.2.2 VALIDACIÓN DE COHERENCIA ENTRE RESULTADOS

Con la ayuda del diagnóstico prescrito por el médico experto se validó la fiabilidad de los resultados obtenidos en la red neuronal. Así mismo, para dar solidez a la validación del resultado obtenido en la red neuronal se comparó con el resultado obtenido en uno de los otros métodos de aprendizaje automático desarrollados.

4.6.2.3 ANÁLISIS DE LOS VARIABLES PREDICTIVAS Y DIAGNÓSTICAS DE LA RED NEURONAL

Luego de validar la coherencia entre los resultados, se realizó una comparativa entre el comportamiento probabilístico y diagnóstico de la red con el ajuste de algunas variables. Todo esto con el fin de determinar la precisión y exactitud de la red si únicamente se alimenta con datos de las variables de mayor influencia para el diagnóstico de la patología.

4.6.2.4 DESARROLLO DE INTERFAZ GRÁFICA

Para el desarrollo de la interfaz se tomaron en consideración las variables que alimentan la base datos con el objetivo de poder analizar datos nuevos, se utilizaron 11 labels para los títulos, 10 textbox y una imagen de fondo. Se determinó que el funcionamiento que se pretende que tenga la interfaz sea que al ingresar los datos se abra una nueva ventana que muestre el diagnóstico, sea positivo o negativo, junto con consejos o recomendaciones según el diagnóstico dado. Así mismo, para los pacientes con probabilidad de padecer diabetes se busca que se abra una ventana diferente donde se revele el porcentaje de probabilidad de padecer la patología junto con consejos y recomendaciones.

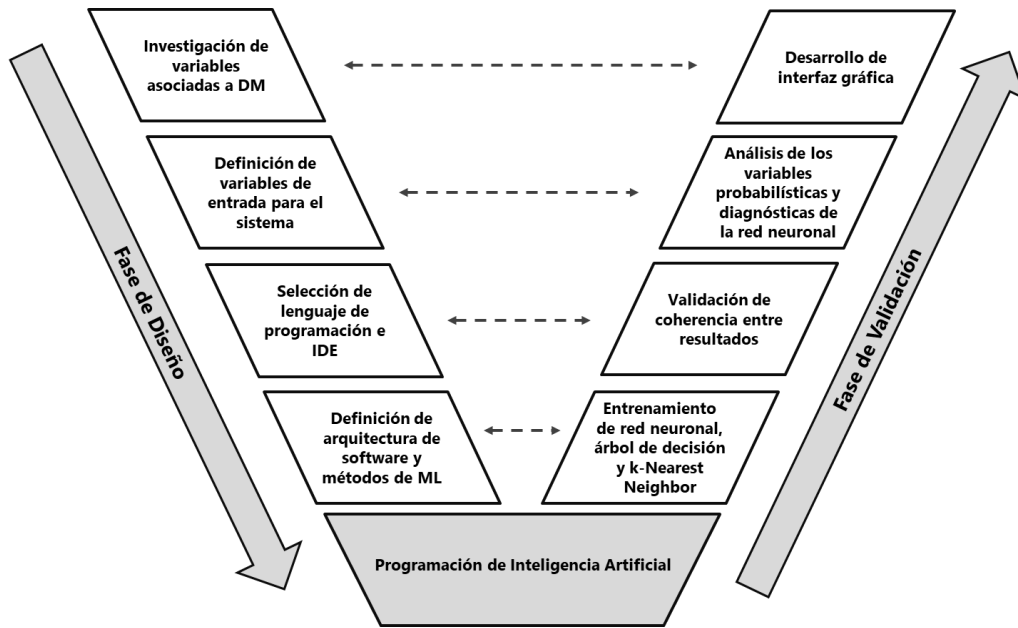


Ilustración 15 Metodología en V.

Fuente: Elaboración propia.

4.7 METODOLOGÍA DE VALIDACIÓN

Con el objetivo de validar el funcionamiento de la red neuronal artificial se utilizaron dos métodos que garantizan que los resultados obtenidos son confiables, para eso se realizó la comparación entre el diagnóstico de experto y el diagnóstico brindado por los métodos utilizados de ML. El diagnóstico profesional es una forma de validación importante ya que un experto de la salud es capaz de utilizar su conocimiento y experiencia para evaluar los síntomas, la historia clínica y los resultados de pruebas para hacer un diagnóstico preciso. Este proceso es útil al momento de detectar cualquier error o incertidumbre en los resultados de la red neuronal y asegurar que el diagnóstico y predicción sean precisos y confiables.

La combinación de un segundo método de aprendizaje automático con el diagnóstico profesional es un enfoque efectivo para validar los resultados de una red neuronal en el campo de la medicina.

4.8 CRONOGRAMA DE ACTIVIDADES

Se realizaron un conjunto de actividades durante un periodo aproximado de 10 semanas para llevar a cabo el proyecto de investigación, todas ellas presentadas por medio de un diagrama de Gantt (Tabla 11).

Tabla 11 Diagrama de Gantt.

Nombre de la tarea	Fecha de inicio	Fecha de finalización	Semana 1	Semana 2	Semana 3	Semana 4	Semana 5	Semana 6	Semana 7	Semana 8	Semana 9	Semana 10
Redacción de carta para solicitud de base de datos.	16/01/2023	19/01/2023	■									
Visita a establecimientos de salud para solicitar base de datos.	24/01/2023	30/01/2023		■	■							
Recopilación de base de datos de entrenamiento y prueba.	30/01/2023	17/02/2023			■	■	■					
Programación de la RNA.	08/02/2023	04/03/2023				■	■	■	■			
Definición de la metodología a utilizar.	10/02/2023	13/02/2023				■	■					
Selección de métodos de inteligencia artificial a utilizar para validación.	14/02/2023	17/02/2023					■					
Elaboración de modelo de árbol de decisión.	19/02/2023	20/02/2023						■				
Elaboración del algoritmo de K-NN.	20/02/2023	21/02/2023						■				
Comparación de resultados obtenidos en los distintos modelos de IA.	22/02/2023	27/02/2023						■	■			
Desarrollo de interfaz gráfica.	28/02/2023	08/03/2023							■	■		
Integración de modelos de IA a la interfaz gráfica.	02/03/2023	14/03/2023							■	■	■	
Elaboración de artículo científico	24/03/2023	26/03/2023										■
Finalización de la investigación	26/03/2023	26/03/2023										■

Fuente: Elaboración propia.

4.9 MATRIZ METODOLÓGICA

La matriz metodológica permite visualizar de forma general la relación de cada uno de los aspectos en ella. En esta se plantea el problema de investigación, preguntas de investigación con sus respectivos objetivos y variables, además de las metodología y herramientas utilizadas para el desarrollo del proyecto (Tabla 12).

Tabla 12 Matriz Metodológica.

Título	Problema de investigación	Preguntas de investigación	Objetivos	Variables	Metodologías y herramientas
		General	General	Dependiente	
Desarrollo de una red neuronal artificial para el diagnóstico clínico y probabilístico de Diabetes Mellitus.	Aunque la glucemia plasmática tiene un gran impacto a la hora de diagnosticar la DM, con ayuda de la inteligencia artificial se puede brindar un apoyo al personal de salud agilizando los procesos con el mínimo porcentaje de error. Por otro lado, la glucemia capilar es el método diagnóstico más sobresaliente debido a la rapidez en la obtención del resultado, sin embargo, este también tiene sus limitaciones	¿Cómo desarrollar un método de diagnóstico para la Diabetes Mellitus con capacidades clínicas probabilísticas utilizando inteligencia artificial?	Desarrollar un método de diagnóstico para la Diabetes Mellitus con capacidades clínicas probabilísticas utilizando inteligencia artificial a través de un algoritmo computacional desarrollado en lenguaje de programación Python.	Red neuronal para el diagnóstico preciso y predicción de DM.	Visual Studio Code Minitab Excel
		Específicas	Específicos	Independientes	
		¿Cuál es el enfoque de aplicabilidad que se le ha dado al aprendizaje automático en el contexto clínico diagnóstico de la Diabetes Mellitus por parte de otros	Recopilar información sobre la aplicación de aprendizaje automático por parte de otros investigadores en el contexto clínico diagnóstico de la Diabetes Mellitus desde el año 2013 hasta la	Variables de entrada de la red.	

que pueden afectar la validez de la lectura.	investigadores a partir del año 2013?	actualidad.	
	¿Cuál es la funcionalidad y aplicabilidad de distintos métodos de inteligencia artificial como herramientas diagnósticas y probabilísticas de la Diabetes Mellitus?	Evaluar la funcionalidad y aplicabilidad de tres métodos de inteligencia artificial como herramientas diagnósticas y probabilísticas de la Diabetes Mellitus.	Modelos de aprendizaje automático.
	¿Cuáles son los principios computacionales y matemáticos que se pueden aplicar en el desarrollo de un red neuronal?	Identificar los principios computacionales y matemáticos que se pueden aplicar en el desarrollo de una red neuronal alimentada con una muestra de datos de pacientes hondureños.	Arquitectura de la red. Datos de entrenamiento.
	¿Qué efecto tiene el uso de librerías durante el desarrollo de una red neuronal para diagnosticar la Diabetes Mellitus y determinar la probabilidad de padecer la patología?	Determinar el efecto que tiene el uso de librerías durante el desarrollo de una red neuronal para diagnosticar la Diabetes Mellitus y determinar la probabilidad de padecer la patología.	

Fuente: Elaboración propia.

4.10 OPERACIONALIZACIÓN DE VARIABLES

La operacionalización de variables ayuda a identificar la relación existente entre las variables y los objetivos de la investigación (Tabla 13). De la misma manera, proporciona información sobre las dimensiones e indicadores de cada variables y una breve descripción de las mismas.

Tabla 13 Operacionalización de Variables.

Objetivo General	Variable Dependiente	Definición Conceptual	Dimensiones	Indicadores
Desarrollar un método de diagnóstico para la Diabetes Mellitus con capacidades clínicas probabilísticas utilizando inteligencia artificial a través de un algoritmo computacional desarrollado en lenguaje de programación Python.	Red neuronal para el diagnóstico preciso y predicción de DM	El desarrollo de un método de diagnóstico para la Diabetes Mellitus con capacidades clínicas predictivas utilizando inteligencia artificial.	Método de diagnóstico. Capacidades clínicas probabilísticas. Inteligencia artificial. Python.	Método de diagnóstico funcional. Probabilidad asertiva. Versión final del prototipo de red neuronal artificial.
Objetivos Específicos	Variable Independiente	Definición Conceptual	Dimensiones	Indicadores
Recopilar información sobre la aplicación de aprendizaje automático por parte de otros investigadores en el contexto clínico diagnóstico de la Diabetes Mellitus desde el año 2013 hasta la actualidad.	VARIABLES DE ENTRADA DE LA RED.	La cantidad de variables o neuronas de entrada utilizadas en la red tienen un gran impacto en el porcentaje de precisión de la misma. El número de variables va a depender directamente de la cantidad de características con la que se quiera entrenar a la red en este caso, resultados de exámenes de laboratorio y preclínica.	Antecedentes de aplicación de aprendizaje automático. Diagnóstico clínico. Diagnóstico probabilístico.	Recopilación y revisión de fuentes bibliográficas.
Evaluar la funcionalidad y aplicabilidad de tres métodos de inteligencia artificial como herramientas diagnósticas y	Modelos de aprendizaje automático.	Existen modelos de aprendizaje automático más precisos que otros según la tarea que se desea realizar. Por tal motivo, diversos métodos de	Inteligencia artificial. Funcionalidad. Aplicabilidad.	Utilidad de la inteligencia artificial en el diagnóstico clínico y

probabilísticas de la Diabetes Mellitus.		aprendizaje automático pueden ser más funcionales que otros para comparar los resultados generados por la red neuronal y evaluar el rendimiento de la misma.	Herramientas diagnósticas y probabilísticas.	probabilístico para Diabetes.
Identificar los principios computacionales y matemáticos que se pueden aplicar en el desarrollo de una red neuronal alimentada con una muestra de datos de pacientes hondureños.	Arquitectura de la red Datos de entrenamiento	La arquitectura de las redes neuronales artificiales se basa en cómo está organizada. Existen diversos tipos de arquitecturas dependiendo de la cantidad de capas, nodos o neuronas y sus características. (Sandoval-Ruiz, 2019).	Organización. Capas. Datos de referencia. Preparación de datos para entrenar a la red.	Tipo de red. Cantidad de capas. Cantidad de datos de referencia. Porcentaje de precisión.
¿Qué efecto tiene el uso de librerías durante el desarrollo de una red neuronal para diagnosticar la Diabetes Mellitus y determinar la probabilidad de padecer la patología?		Las redes neuronales artificiales deben ser alimentadas con una amplia base de datos de entrenamiento para tomar como punto de referencia para el aprendizaje. Mientras mayor sea la base de datos con la que se alimente a la red, mayor será la cantidad de información que esta aprende y, por lo tanto, menor será el porcentaje de error generado.		

Fuente: Elaboración propia

V. RESULTADOS Y ANÁLISIS

Siguiendo la estructura de la metodología en V, este capítulo expone los resultados obtenidos en cada una de las etapas, tanto de la fase de diseño como de la fase de validación. Incluyendo también otros resultados que proporcionan información de gran valor para la investigación.

5.1 RELACIÓN DE VARIABLES DE ENTRADA PARA EL SISTEMA

El mapa de calor representa gráficamente los valores en una matriz de datos. Esta técnica de visualización de datos permite realizar un análisis de relación de variables mediante una gama de colores. Se realizó el mapa de calor tomando en consideración las variables de entrada utilizadas en los modelos de aprendizaje automático para realizar el diagnóstico y probabilidad de padecer diabetes mellitus.

En la comparación de cada una de estas variables de entrada con el Target, que es el que indica si la persona padece o no DM, se identificó que aquellas variables con mayor influencia sobre el diagnóstico son la glucosa en sangre y la hemoglobina glicosilada (Ilustración 16) Lo cual coincide con el criterio de los médicos y lo establecido según la Asociación Americana de Diabetes.

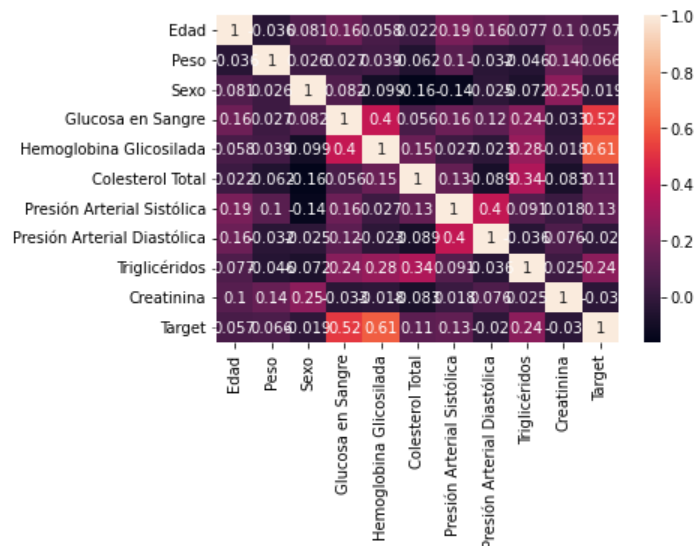


Ilustración 16 Mapa de calor de análisis de grupos.

Fuente: Elaboración propia.

Debido a la importancia en las variables "glucosa en sangre" y "hemoglobina glicosilada" para la probabilidad y diagnóstico de DM, se realizó un diagrama de dispersión de dichas variables evaluándolo desde ambos casos, el paciente diabético y aquel sin diagnóstico de DM. Tomando en el eje "y" los valores de glucosa en sangre y en el eje "x" los registros, en el caso de pacientes diabéticos se tienen 105 registros y en pacientes sin diagnóstico de DM un total de 65 registros.

En los pacientes sin un diagnóstico de diabetes de la base de datos recopilada los valores de glucosa en sangre se encontraron concentrados principalmente entre 68 y 122, a excepción de un valor levemente atípico donde la glucosa en sangre fue de 138 (Ilustración 17).

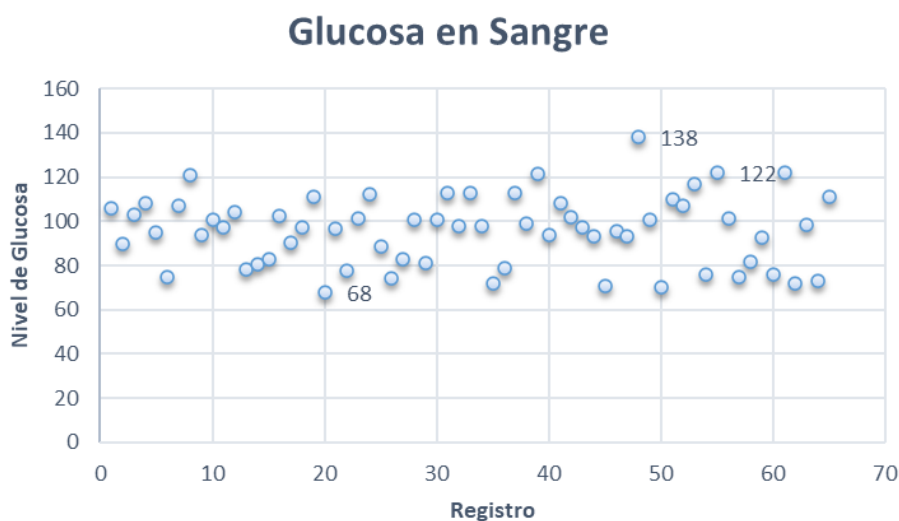


Ilustración 17 Diagrama de dispersión de valores de glucosa en sangre en pacientes sin diagnóstico de DM.

Fuente: Elaboración propia.

Por otro lado, en el diagrama de dispersión de la glucosa en sangre en pacientes diabéticos estos valores de glucosa se encontraban ubicados entre 66 y 475 (Ilustración 18). De igual manera, se tuvo un valor atípico extremo de 1023 que, aunque no suele ser tan común, si ha habido casos de pacientes diabéticos con esos valores de glucosa.

Glucosa en Sangre

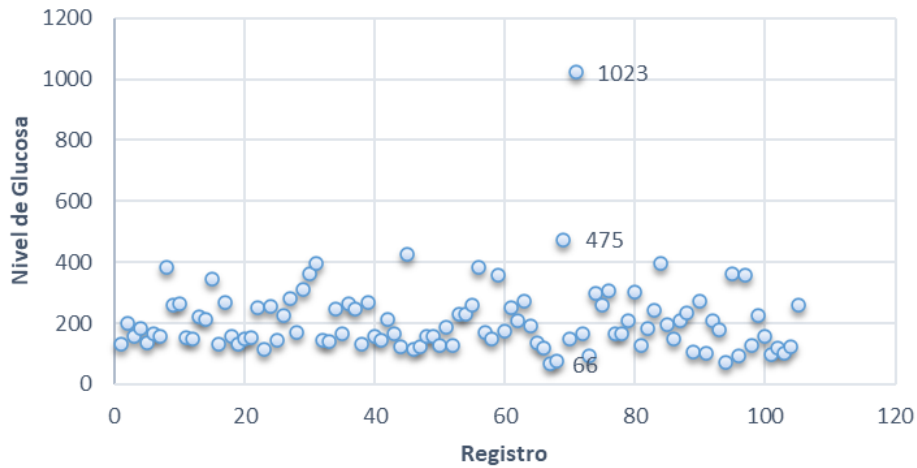


Ilustración 18 Diagrama de dispersión de valores de glucosa en sangre en pacientes diabéticos.

Fuente: Elaboración propia.

En cuanto a la hemoglobina glicosilada en pacientes sin diagnóstico de DM, el diagrama de dispersión mostró que este tipo de pacientes, en su mayoría, obtuvieron resultados entre 3.8 y 6.9 (Ilustración 19).

Hemoglobina Glicosilada

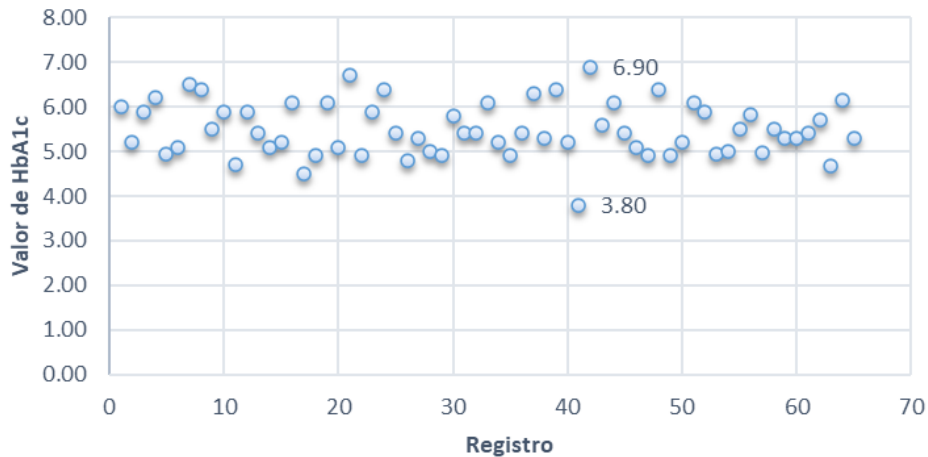


Ilustración 19 Diagrama de dispersión de valores de hemoglobina glicosilada en pacientes sin diagnóstico de DM.

Fuente: Elaboración propia.

En cambio, en pacientes diabéticos, los valores de hemoglobina glicosilada obtenidos se encontraban entre los 5.50 y 15.40, sobre todo entre 5.50 y 9.40 (Ilustración 20).

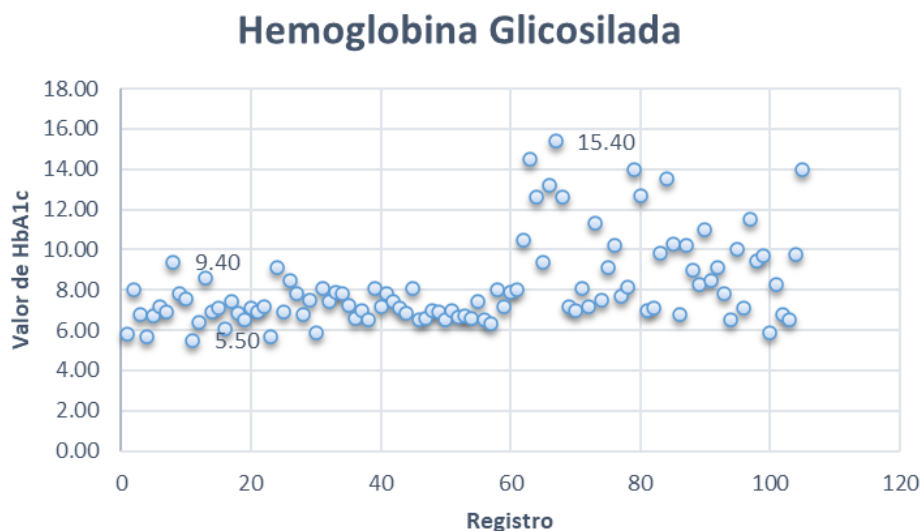


Ilustración 20 Diagrama de dispersión de valores de hemoglobina glicosilada en pacientes diabéticos.

Fuente: Elaboración propia.

5.2 ENTRENAMIENTO DE MÉTODOS DE APRENDIZAJE AUTOMÁTICO

Para lograr que los resultados brindados por los distintos métodos de aprendizaje automático sean lo más acertados posible se debe acudir al entrenamiento de los mismos. Cada uno de ellos con una forma distinta de ser entrenados hasta que se obtengan los resultados deseados con el menor error probable.

5.2.1 RED NEURONAL ARTIFICIAL CON KERAS Y TENSORFLOW

Con el objetivo de identificar la fiabilidad y precisión de la red se graficó el comportamiento de la pérdida durante los ciclos definidos (epochs), también conocidos como épocas. El número de ciclos definidos representa la cantidad de veces que se ejecutó el método de backpropagation, o retropropagación, que es mediante el cual se entrena a la red.

La función de pérdida en una red neuronal indica qué tan bien está funcionando la red. Una función de pérdida elevada indica que la red tiene un rendimiento bastante pobre, por ello se debe disminuir utilizando el método de backpropagation.

En el caso de datos conocidos cuya gráfica se visualiza en color azul, es decir la pérdida de entrenamiento, mientras las épocas avanzan, las pérdidas disminuyen (Ilustración 21). Durante el proceso de entrenamiento se ajustó el resultado arrojado por la red variando la cantidad de ciclos, de modo que las pérdidas fueron casi 0, lo que garantiza que el resultado es confiable.

Por otro lado, la pérdida en los datos de prueba, o pérdida de validación, se ve representada en color naranja en la gráfica y se visualiza que inicia con un buen ajuste ya que disminuye, pero a partir de los 100 ciclos aproximadamente se mantiene casi en el mismo valor de pérdida (Ilustración 21). Al ser mayor la pérdida de validación que la pérdida de entrenamiento, se dice que existe un sobreajuste en el modelo lo que significa que ha sido entrenado por mucho tiempo o que la base de datos es muy pequeña. Este último es el caso de este modelo de red neuronal.

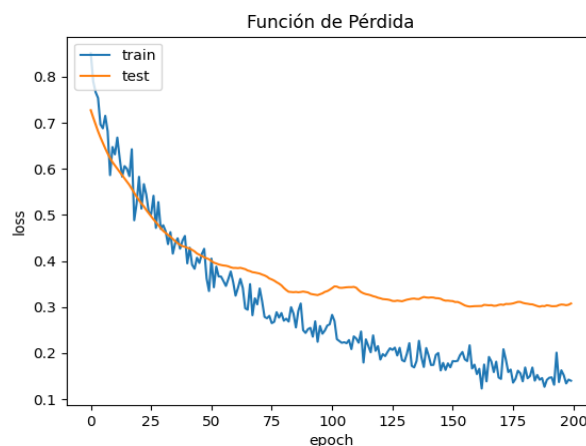


Ilustración 21 Función de pérdida con respecto a las épocas en la red neuronal con Keras y TensorFlow.

Fuente: Elaboración propia.

La precisión del modelo de red neuronal desarrollado demostró ser de 94% para datos de entrenamiento, es decir, datos conocidos, y de 91% para datos de prueba (Ilustración 22). De los anteriores planteamiento se deduce que el modelo funciona mucho mejor con los datos con los que ha sido entrenado que para aquellos donde requiere realizar predicciones. Sin embargo,

91% es un porcentaje lo suficientemente alto para decir que la red neuronal realiza un diagnóstico clínico y probabilístico bastante confiable utilizando datos nuevos.

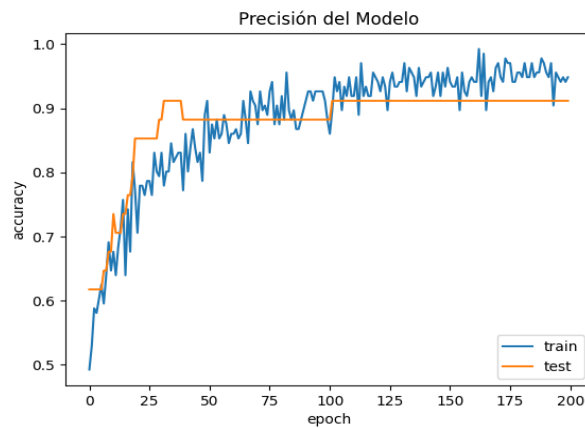


Ilustración 22 Precisión del modelo de la red neuronal con Keras y TensorFlow.

Fuente: Elaboración propia.

Para el desarrollo del prototipo se llevó a cabo una secuencia lógica específica apoyada con la arquitectura de las librerías TensorFlow y Keras (Ilustración 23). La inicialización del código se basó en la carga de las librerías requeridas seguida de los siguientes pasos:

1. Carga de datos: En primer lugar, se cargaron los datos de entrenamiento y prueba de un conjunto de datos con información sobre pacientes, incluyendo factores como edad, nivel de glucosa en sangre, colesterol, etc.
2. Clasificación de datos de prueba y entrenamiento: El código generó una distribución aleatoria de datos separando en 80% para datos de entrenamiento y 20% para datos de prueba. Teniendo en total 136 registros para el entrenamiento y 34 para prueba.
3. Normalización de datos: Antes de alimentar los datos en la red neuronal, es común normalizarlos para asegurarse de que todas las características tengan una escala similar y estén centradas en cero. Esto ayuda a que la red neuronal aprenda de manera más efectiva.
4. Definición del modelo de red neuronal: El modelo de red neuronal se diseñó bajo la estructura de 3 capas. La primera con 32 neuronas y una función de activación "relu", la segunda es una capa dropout y se aplicó con el fin de no sobre ajustar el entrenamiento

de la red, por último, una capa densa de salida con una única neurona y una función de activación "sigmoidea".

5. Compilación del modelo: Una vez definido el modelo, se compiló con una función de pérdida adecuada (en este caso, `binary_crossentropy`) para medir la discrepancia entre la salida real y la predicción del modelo, un optimizador (en este caso, Adam) para ajustar los pesos del modelo durante el entrenamiento y una métrica de precisión "accuracy" para medir el rendimiento del modelo.
6. Entrenamiento del modelo: Con el modelo compilado, se pudo entrenar utilizando el conjunto de datos de entrenamiento normalizado. El modelo se entrenó a lo largo de varias épocas (iteraciones), en este caso 200 y se actualizó utilizando el método de retro propagación con el fin de ajustar los pesos de las neuronas en la red. Durante el entrenamiento, se utilizó una fracción de los datos de entrenamiento para validar el modelo y asegurarse de que no se está sobre ajustando.
7. Guardar el modelo entrenado: Una vez entrenada la red, se guardaron los datos y relaciones que se consiguieron, preparando la red para el análisis de nuevos datos.
8. Evaluación del funcionamiento del modelo: Una vez que se completó el entrenamiento y después de haberse guardado, se evaluó el modelo utilizando el conjunto de datos de prueba. Esto proporcionó una idea de cómo se desempeñó el modelo en datos no vistos.
9. Ingresar datos de prueba: Con el fin de obtener un resultado, se ingresaron datos de nuevos pacientes para ser analizados por la red correspondientes a cada una de las variables estudiadas.
10. ¿Cuál es el diagnóstico? Al momento de la evaluación, la red presentó 3 tipos de respuesta diferentes; positivo, negativo probabilístico y negativo. La pantalla que se muestra depende del diagnóstico de la red. Cada pantalla mostrada cuenta con recomendaciones pertinentes a cada diagnóstico.

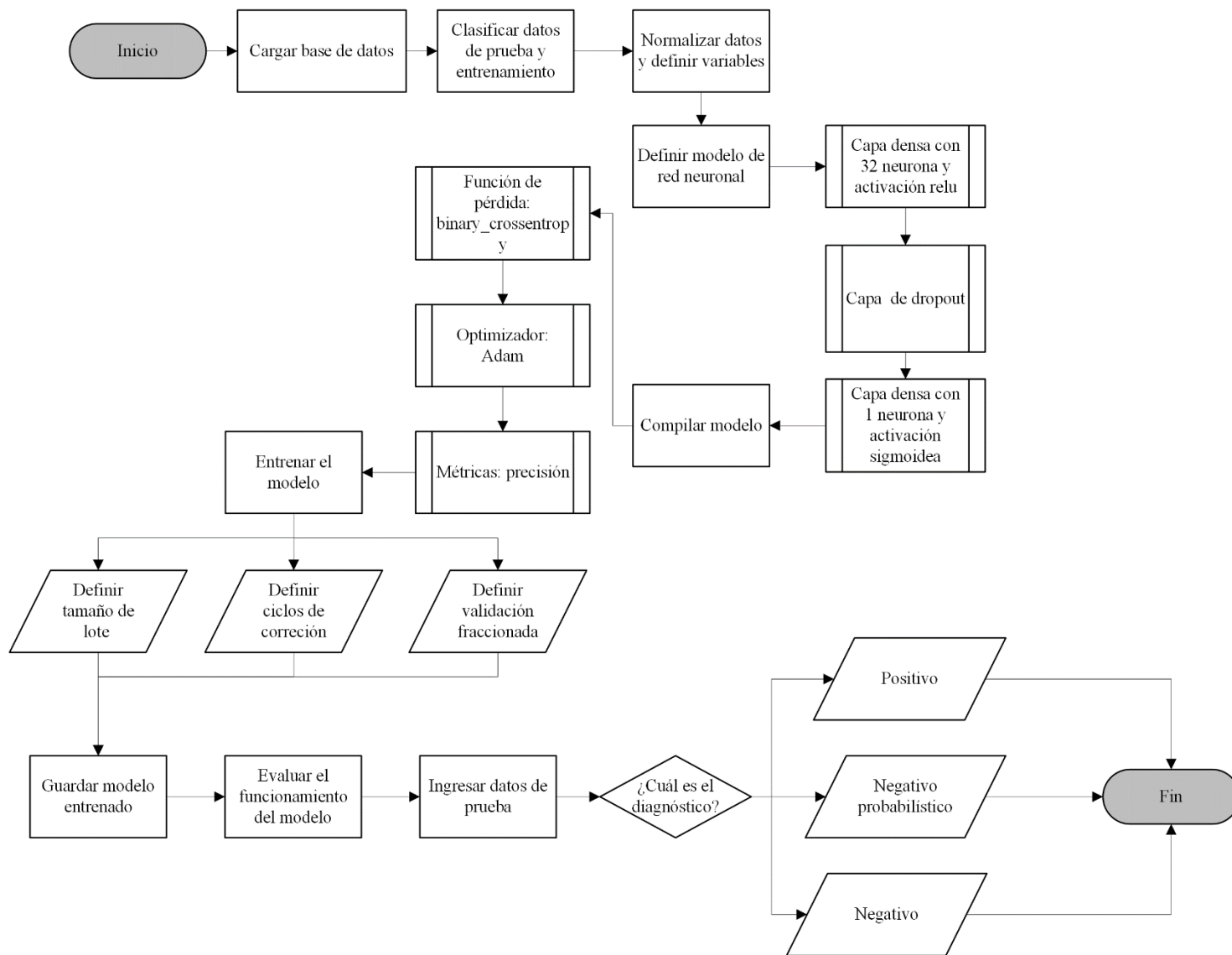


Ilustración 23 Funcionamiento lógico de la red neuronal con TF y Keras.

Fuente: Elaboración propia.

5.2.2 K-NEAREST NEIGHBOR ALGORITHM

Se realizó una gráfica para visualizar el error generado para la clasificación en el algoritmo del método de aprendizaje automático K-Nearest Neighbor. En esta se representó en el eje "y" el error generado y en el eje "x" el valor de k utilizado. El valor de k representa la cantidad de vecinos que fueron verificados para realizar la clasificación.

La gráfica mostró que el error está en su punto mínimo cuando el valor de k se encontraba entre k=2 y k=4 aproximadamente (Ilustración 24). Sin embargo, según la programación realizada, el modelo trabaja de mejor manera cuando k=8.

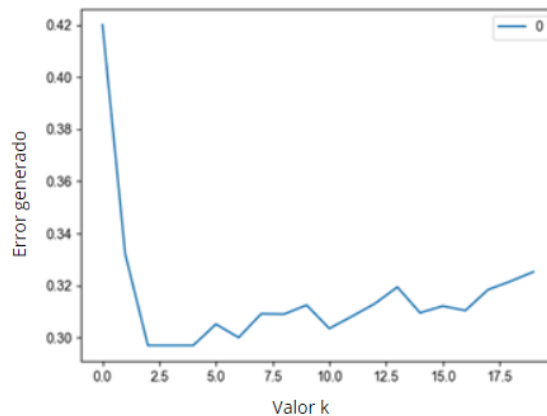


Ilustración 24 Error en el modelo de k-NN con respecto al k utilizado.

Fuente: Elaboración propia.

5.2.3 ÁRBOL DE DECISIÓN

El modelo de árbol de decisión utilizado se realizó mediante la librería scikit-learn de Python, la cual utilizó la base de datos de entrenamiento de 136 registros para crear un modelo que sea eficiente y preciso para la clasificación binaria deseada.

Este trabaja realizando comparaciones de valores numéricos para clasificar como clase 0 a aquellos pacientes sin diagnóstico de DM y como clase 1 a los pacientes diabéticos. Inicialmente, el modelo considera un valor de glucosa en sangre menor o igual a 122.5, de ser verdadera esta condición, el modelo continúa evaluando los parámetros introducidos hasta

llegar a tomar una decisión en cuanto a la clasificación. Por otro lado, de ser falso, inmediatamente lo clasifica como clase 1, es decir, diabético (Ilustración 25).

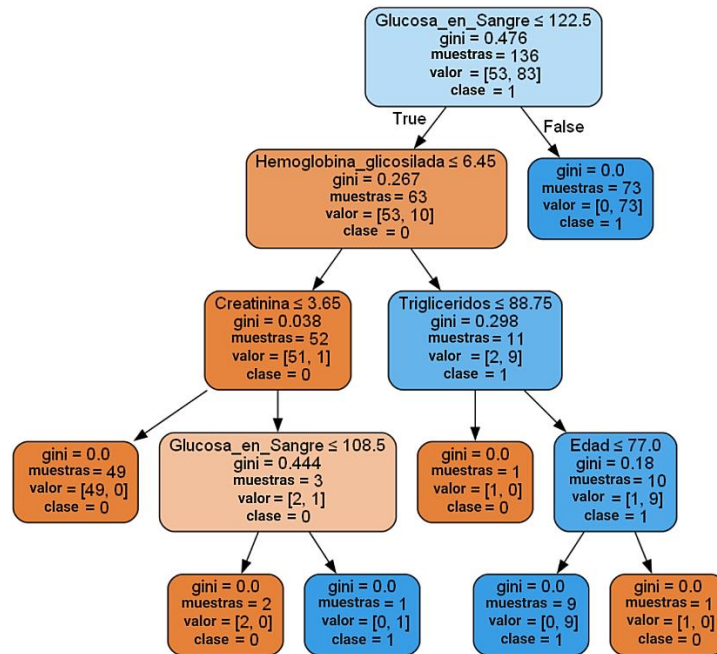


Ilustración 25 Modelo de árbol de decisión.

Fuente: Elaboración propia.

5.3 VALIDACIÓN DE COHERENCIA ENTRE RESULTADOS

Con el fin de determinar la precisión y utilidad de los distintos métodos de inteligencia artificial, se acudió al uso de herramientas como gráficos, matrices de confusión, reportes de la clasificación brindada por los modelos y comparaciones entre los resultados obtenidos en cada uno de ellos con el diagnóstico dado por el experto.

5.3.1 ÁRBOL DE DECISIÓN

Se utilizó una matriz de confusión para evaluar el rendimiento del algoritmo de árbol de decisión. Por medio de esta se determinan las predicciones o clasificaciones realizadas por el modelo que fueron acertadas o desacertadas.

En la esquina superior izquierda se indican los valores verdaderos negativos, es decir, donde el modelo predice correctamente la clase negativa. La esquina superior derecha muestra

los falsos positivos, conocido como error de tipo 1, donde el modelo clasifica como positivos a aquellos que en realidad son negativos.

El error de tipo 2 se ve representado por los falsos negativos en la esquina inferior izquierda de la matriz de confusión, este es donde el modelo clasifica erróneamente a la clase positiva como clase negativa. Y, finalmente, en la esquina inferior derecha se realiza un conteo de aquellos positivos clasificados correctamente por el modelo, llamados verdaderos positivos.

Según la base de datos de prueba (Anexo 3), la matriz de confusión realizó un conteo donde se obtuvieron resultados significativos para evaluar el rendimiento del modelo. Con un total de 10 resultados verdaderos negativos, 22 verdaderos positivos, 0 falsos negativos y 2 falsos positivos (Ilustración 26).

En referencia a la clasificación anterior, se concluye que el modelo de árbol de decisión realiza una clasificación de resultado negativo bastante precisa, ya que no erró en ninguna ocasión. Mientras que en la clasificación realizada para resultado positivo el modelo acertó en 22 de los 24 casos, lo cual, de igual manera indica que es bastante preciso.

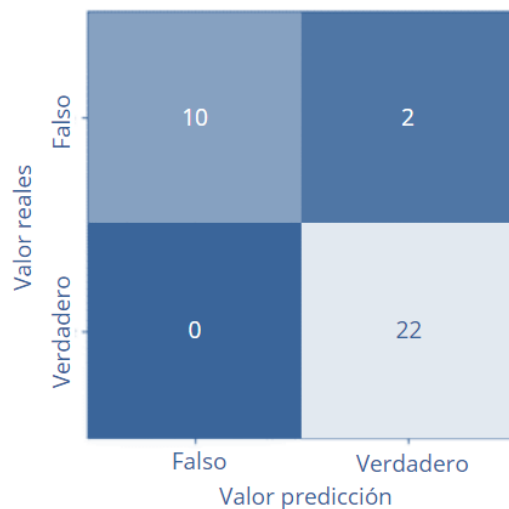


Ilustración 26 Matriz de confusión según la clasificación realizada por el modelo de árbol de decisión.

Fuente: Elaboración propia.

Lo anterior expuesto se puede visualizar también en forma de porcentajes en un reporte de la clasificación realizada por el algoritmo de árbol de decisión donde se determina el

porcentaje de precisión para clasificar los negativos y los positivos (Tabla 14). El modelo demostró tener una precisión para la clasificación de resultado negativo de 100% y, en el caso de resultado positivo un 92%.

La exhaustividad, al igual que la precisión, son criterios importantes para evaluar la efectividad del modelo. Mientras la precisión mide la calidad de las predicciones realizadas por el modelo, la exhaustividad se refiere a la capacidad del modelo para clasificar todas las muestras de los datos de prueba correctamente. Como ejemplo, este mide la proporción de casos positivos correctamente identificados en comparación con todos los casos positivos reales existentes en los datos de prueba.

La puntuación-F1 brinda información relevante cuando se quiere tener un valor más objetivo ya que compara el rendimiento combinado de la precisión y exhaustividad del modelo, en este caso, para la clasificación de diagnóstico negativo se obtuvo un 91% y para el diagnóstico positivo un 96%.

Tabla 14 Reporte de clasificación del modelo de árbol de decisión.

Clasificación	Precisión	Exhaustividad	Puntuación – F1	Cantidades
Diagnóstico Negativo	1.00	0.83	0.91	12
Diagnóstico Positivo	0.92	1.00	0.96	22

Fuente: Elaboración propia.

Finalmente, el modelo de árbol de decisión demostró tener una precisión total de 94.11% para la clasificación de diagnóstico positivo y negativo de diabetes mellitus.

5.3.2 K-NEAREST NEIGHBOR

Se introdujeron al algoritmo los datos de entrenamiento y se consideraron las variables con mayor influencia para el diagnóstico de DM para realizar los gráficos, visualizando la glucosa en sangre en el eje "x" y la hemoglobina glicosilada en el eje "y". Al graficar esto, el algoritmo realizó una agrupación de los datos e identificó a los pacientes diabéticos en color negro y a los pacientes sin diagnóstico de DM en color blanco.

En tal agrupación se logró observar que, en el conjunto de datos analizados, es decir, la base de datos de entrenamiento, los pacientes sin diagnóstico de DM se veían más orientados

en dirección hacia la esquina inferior izquierda del plano, mientras que los pacientes diabéticos se encontraban más dispersos (Ilustración 27).

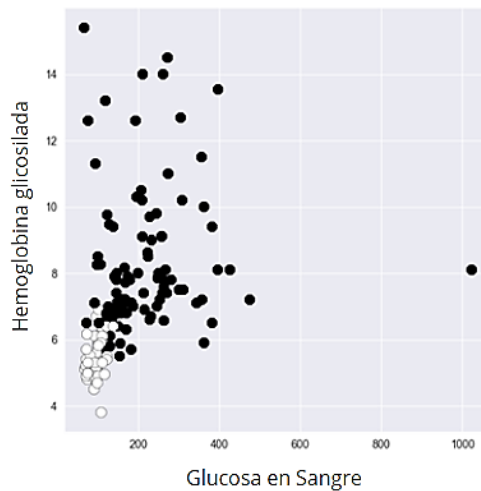


Ilustración 27 Clasificación de pacientes según el modelo kNN utilizando la base de datos de entrenamiento.

Fuente: Elaboración propia.

Considerando el valor de k con mejor rendimiento según la programación realizada, $k=8$, y otro valor de k donde el error es mayor, como $k=1$, se realizó una comparación para identificar las diferencias en ambos modelos utilizando la base de datos de prueba (Anexo 3) para evaluar las características predictivas del modelo (Ilustración 28).

Ya que previamente se evaluó el algoritmo con la base de datos de entrenamiento y se visualizó de forma general cual era la predisposición en cuanto a la ubicación de los datos según su clasificación, se puede tener una idea de cómo los datos de prueba deberían ubicarse también según su clasificación.

Con $k=1$ se observó que en la esquina inferior izquierda se ubicaban tanto pacientes diabéticos como pacientes sin diagnóstico de DM, mientras que cuando se utilizó la base de datos de entrenamiento únicamente se encontraban pacientes sin diagnóstico de DM. Este modelo, utilizando $k=1$ resultó tener una precisión de 82.35% para la clasificación de datos nuevos.

Por otro lado, cuando se utilizó $k=8$ con los mismos datos de prueba, se contempló una notable diferencia en cuanto a la clasificación de diagnóstico de DM realizada. Clasificaciones categorizadas con color negro utilizando $k=1$ en realidad debieron haber estado en color blanco. Mientras que al utilizar $k=8$, el gráfico se mostró con una distribución más parecida a la brindada cuando se utilizó la base de datos de entrenamiento. La precisión del modelo con $k=8$ fue de 88.23%.

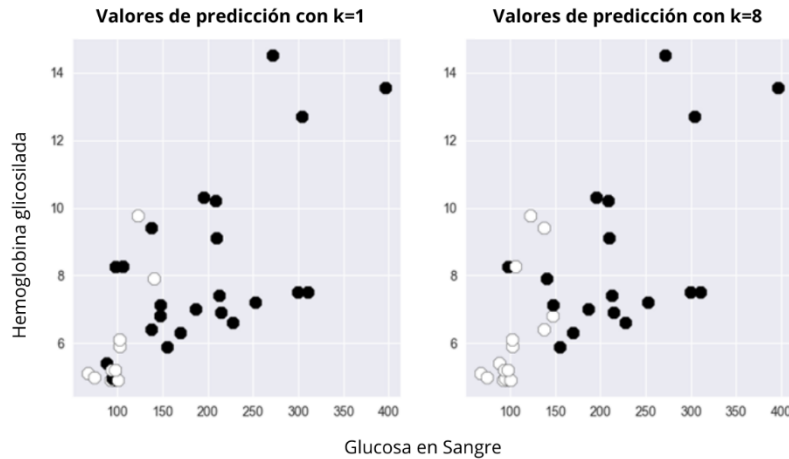


Ilustración 28 Agrupación de valores de predicción por el modelo kNN utilizando valores de k diferentes.

Fuente: Elaboración propia.

5.3.3 RED NEURONAL CON KERAS Y TENSORFLOW

Con el fin de evaluar el rendimiento de la red neuronal se realizó una comparación entre los resultados obtenidos por la red, los obtenidos mediante el algoritmo de árbol de decisión, debido a que este obtuvo mejores resultados que el algoritmo de k-nearest neighbor, y el diagnóstico dado por el experto (Tabla 15).

Se tomaron los 34 registros de la base de datos de prueba (Anexo 3) y se ingresaron manualmente tanto en el modelo de red neuronal como en el del árbol de decisión. Luego, se realizó una comparación de los diagnósticos brindados por la red para realizar un conteo de aquellos que coincidían tanto con el resultado brindado por el árbol de decisión como por el brindado por el médico experto.

Cabe resaltar que, tanto la red neuronal como el árbol de decisión realizan la clasificación de diagnóstico como positivo o negativo. Sin embargo, en aquellos diagnósticos negativos que la red neuronal llega a considerar como pacientes prediabéticos o posibles diabéticos se les brinda también, además de la clasificación, el porcentaje de probabilidad de padecer DM, cosa que el árbol de decisión no proporciona.

Tabla 15 Validación resultados obtenidos en la red neuronal.

Número de registro	Resultado red neuronal	Resultado árbol de decisión	Diagnóstico de experto
1	Negativo (75.96%)	Positivo	Positivo
2	Negativo	Negativo	Negativo
3	Positivo	Positivo	Positivo
4	Negativo (22.71%)	Positivo	Negativo
5	Positivo	Positivo	Positivo
6	Positivo	Positivo	Positivo
7	Positivo	Positivo	Positivo
8	Negativo (32.33%)	Positivo	Negativo
9	Positivo	Positivo	Positivo
10	Positivo	Positivo	Positivo
11	Positivo	Positivo	Positivo
12	Positivo	Positivo	Positivo
13	Positivo	Positivo	Positivo
14	Negativo (72.22%)	Positivo	Positivo
15	Negativo (22.25%)	Negativo	Negativo
16	Positivo	Positivo	Positivo
17	Negativo	Negativo	Negativo
18	Negativo	Negativo	Negativo
19	Positivo	Positivo	Positivo
20	Positivo	Positivo	Positivo
21	Positivo	Positivo	Positivo
22	Positivo	Positivo	Positivo
23	Positivo	Positivo	Positivo
24	Positivo	Positivo	Positivo
25	Positivo	Positivo	Positivo
26	Negativo (57.54%)	Positivo	Positivo
27	Negativo (74.16%)	Positivo	Positivo
28	Negativo	Negativo	Negativo
29	Positivo	Positivo	Positivo
30	Positivo	Positivo	Positivo
31	Positivo	Positivo	Positivo
32	Positivo	Negativo	Negativo
33	Negativo (4.83%)	Negativo	Negativo
34	Negativo (6.26%)	Negativo	Negativo

Fuente: Elaboración propia.

En comparación con el modelo de árbol de decisión, se concluye que la red neuronal tuvo un desempeño bastante similar brindando el mismo resultado en 20 ocasiones para casos positivos y en 7 ocasiones para casos negativos de DM (Tabla 16). Lo que quiere decir que, en total fueron 27 las coincidencias entre los resultados brindados por estos modelos de aprendizaje automático. De los 34 registros de prueba estos modelos concordaron en 27 casos.

Por otro lado, en comparación con el diagnóstico del experto, tanto la red neuronal como el árbol de decisión tuvieron la misma cantidad de concordancias con respecto a la clasificación de diagnóstico positivo, pero con respecto a la clasificación de diagnóstico negativo fue el modelo de árbol de decisión el que tuvo una mayor cantidad de coincidencias (Tabla 16).

En total, fueron 27 las coincidencias entre los resultados brindados por la red neuronal y los brindados por el árbol de decisión. Por otro lado, comparando ambos métodos de aprendizaje automático con el diagnóstico del experto, la red neuronal tuvo un total de 29 coincidencias lo que quiere decir que únicamente erró en 5 ocasiones y el árbol de decisión en 2 ocasiones, ya que obtuvo 32 coincidencias.

Aunque el árbol de decisión obtuvo mayor concordancia con los resultados brindados por el experto, tiene la desventaja de no proporcionar el porcentaje de probabilidad de padecer la patología. Y, aunque la red obtuvo solo 29 coincidencias en comparación con el diagnóstico del experto, muchos de los casos falsos positivos brindados por la red tenían un porcentaje bastante alto de probabilidad de padecer la patología.

Tabla 16 Comparación de resultados obtenidos.

	Concordancia de resultados RNA vs. Árbol de decisión	Concordancia de resultados RNA vs. Diagnóstico de experto	Concordancia de resultados Árbol de decisión vs. Diagnóstico de experto
Positivos	20	20	24
Negativos	7	9	8

Fuente: Elaboración propia.

5.4 ANÁLISIS DE LOS VARIABLES PREDICTIVAS Y DIAGNÓSTICAS

Con la finalidad de identificar si existe alguna diferencia en la precisión de los resultados obtenidos por la red neuronal si se manipulan las variables de entrada, se realizó un segundo modelo para elaborar una comparación. Este segundo modelo de red neuronal utilizó únicamente las 2 variables de mayor influencia en el diagnóstico según el mapa de calor, glucosa en sangre y hemoglobina glicosilada.

Se tomó como Modelo 1 la red neuronal que evalúa todas las variables (edad, peso, sexo, glucosa en sangre, hemoglobina glicosilada, colesterol total, presión arterial sistólica y diastólica, triglicéridos y creatinina) para realizar el diagnóstico clínico y probabilístico. Mientras que, se nombró como Modelo 2 aquel que solamente realiza el análisis con las variables de glucosa en sangre y hemoglobina glicosilada.

Una vez realizado el modelo, se introdujeron los valores de la base de datos de prueba (Anexo 3) y se compararon los resultados obtenidos previamente en el modelo 1, y, de igual manera, el diagnóstico del experto (Tabla 17).

Tabla 17 Comparación de modelos de red neuronal alterando las entradas.

Número de registro	Resultados Modelo 1	Resultados Modelo 2	Diagnóstico de experto
1	Negativo (75.96%)	Negativo (69.99%)	Positivo
2	Negativo	Negativo	Negativo
3	Positivo	Positivo	Positivo
4	Negativo (22.71%)	Negativo (42.53%)	Negativo
5	Positivo	Negativo (48.11%)	Positivo
6	Positivo	Positivo	Positivo
7	Positivo	Positivo	Positivo
8	Negativo (32.33%)	Negativo (49.30%)	Negativo
9	Positivo	Positivo	Positivo
10	Positivo	Positivo	Positivo
11	Positivo	Positivo	Positivo
12	Positivo	Positivo	Positivo
13	Positivo	Positivo	Positivo
14	Negativo (72.22%)	Positivo	Positivo
15	Negativo (22.25%)	Negativo (15.15%)	Negativo
16	Positivo	Positivo	Positivo
17	Negativo	Negativo	Negativo
18	Negativo	Negativo	Negativo
19	Positivo	Negativo (77.71%)	Positivo
20	Positivo	Positivo	Positivo

Continuación Tabla 17

Número de registro	Resultados Modelo 1	Resultados Modelo 2	Diagnóstico de experto
21	Positivo	Positivo	Positivo
22	Positivo	Positivo	Positivo
23	Positivo	Positivo	Positivo
24	Positivo	Positivo	Positivo
25	Positivo	Positivo	Positivo
26	Negativo (57.54%)	Positivo	Positivo
27	Negativo (74.16%)	Negativo (76.07%)	Positivo
28	Negativo	Negativo	Negativo
29	Positivo	Positivo	Positivo
30	Positivo	Positivo	Positivo
31	Positivo	Positivo	Positivo
32	Positivo	Negativo	Negativo
33	Negativo (4.83%)	Negativo (6.86%)	Negativo
34	Negativo (6.26%)	Negativo (8.65%)	Negativo

Fuente: Elaboración propia.

Con referencia a lo anterior, se puede decir que el Modelo 1 y el Modelo 2 tienen capacidades de clasificación similares ya que, en comparación con el diagnóstico del experto, el Modelo 1 erró en 5 ocasiones y el Modelo 2 en 4 (Tabla 18).

Y en cuanto a las capacidades para brindar un diagnóstico probabilístico, los resultados brindados por ambos modelos aparentan ser bastante coherentes por lo que se puede decir que ambos modelos tienen un rendimiento bastante similar a pesar de utilizar distintas variables de entrada.

Tabla 18 Comparación de resultados obtenidos.

	Concordancia de resultados Modelo 1 vs. Modelo 2	Concordancia de resultados Modelo 1 vs. Diagnóstico de experto	Concordancia de resultados Modelo 2 vs. Diagnóstico de experto
Positivos	18	20	20
Negativos	11	9	10

Fuente: Elaboración propia.

5.5 DESARROLLO DE INTERFAZ GRÁFICA

Para facilitar el ingreso de los datos a analizar se desarrolló una interfaz amigable con el usuario, esta cuenta con un cuadro de texto para la entrada de cada una de las variables definidas (Ilustración 29). El prototipo también cuenta con los logos de representativos de la universidad, el de la carrera de Ingeniería en Biomédica, el de la Facultad de Ingeniería y el de la universidad.

Cada uno de los campos se enlazó con la base de datos utilizada. Con el fin de mantener la estética y la armonía visual se abreviaron algunas variables como ser hemoglobina glicosilada (HbaA1c), presión arterial sistólica (PAS) y presión arterial diastólica (PAD).

El funcionamiento de la interfaz es sencillo, se ingresan los valores numéricos en cada uno de los campos y al dar clic en el botón calcular aparece una ventana emergente con el diagnóstico (positivo o negativo). En caso de ser negativo, sobre todo en pacientes que la red considera prediabéticos, la ventana muestra el probabilidad en porcentaje que tiene el paciente de padecer la enfermedad.



Ilustración 29 Interfaz gráfica desarrollada en Visual Studio Code.

Fuente: Elaboración propia.

Junto con lo anterior, se muestra una serie de recomendaciones para cada uno de los diagnósticos: positivo (Ilustración 30), negativo (Ilustración 31) y negativo probabilístico

(Ilustración 32). Por último, al presionar el botón borrar se limpian todos los campos de la interfaz para poder ingresar nuevos valores.

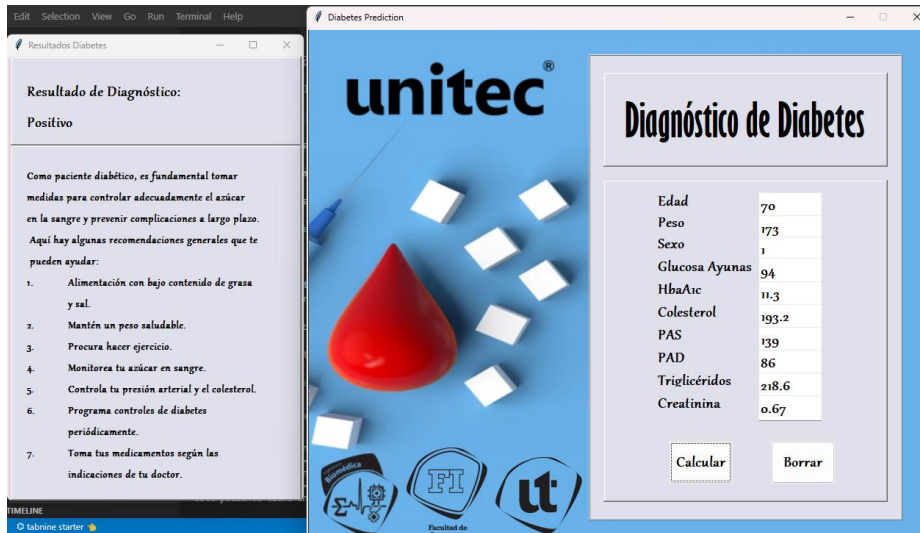


Ilustración 30 Interfaz gráfica, resultado de diagnóstico positivo.

Fuente: Elaboración propia.



Ilustración 31 Interfaz gráfica, resultado de diagnóstico negativo.

Fuente: Elaboración propia.



Ilustración 32 32 Interfaz gráfica, resultado de diagnóstico negativo probabilístico.

Fuente: Elaboración propia.

5.6 ANÁLISIS ESTADÍSTICO DESCRIPTIVO

Durante el desarrollo de la investigación se trabajó con un total de 170 expedientes de pacientes que alimentaron la base de datos. Con el fin de estudiar el comportamiento de cada una de las variables que la componen se acudió a la estadística descriptiva aplicada en minitab. Se analizaron los datos dividiendo los pacientes diabéticos y no diabéticos para poder observar el comportamiento de los mismos y profundizar en conclusiones de valor para la investigación.

Los parámetros estadísticos para evaluar fueron: media, error estándar de la media, desviación estándar, varianza, coeficiente de variación, mínimo, máximo, rango, N valores presentes, N valores faltantes, N acumulado, porcentaje, porcentaje acumulado, primer cuartil, mediana, tercer cuartil, rango Inter cuartil, moda, asimetría y curtosis. Estos parámetros se evaluaron para ambos casos de pacientes, sin diagnóstico de DM (Tabla 19) y con diagnóstico positivo de DM (Tabla 20).

Tabla 19 Estadística descriptiva asociada a pacientes sin diagnóstico de DM.

Variable	Conteo total	N	N*	NAcum	Porcentaje	PrcAcum	Media	Error estándar de la media
Edad	65	65	0	65	100	100	54.00	2.34
Peso	65	65	0	65	100	100	160.92	3.92
Glucosa_en_Sangre	65	65	0	65	100	100	95.89	1.94

Variable	Conteo total	N	N*	NAcum	Porcentaje	PrcAcum	Media	Error estándar de la media
Hemoglobina_glicosilada	65	65	0	65	100	100	5.4962	0.0742
Colesterol	65	65	0	65	100	100	188.12	6.71
Presion_arterial_sistolica	65	65	0	65	100	100	130.60	2.89
Presion_arterial_diastolica	65	65	0	65	100	100	80.75	1.99
Trigliceridos	65	65	0	65	100	100	156.15	8.94
Creatinina	65	65	0	65	100	100	1.109	0.115

Variable	Desv.Est.	Varianza	Mínimo	Q1	Mediana	Q3	Máximo
Edad	18.90	357.16	14.00	41.00	55.00	66.50	91.00
Peso	31.61	999.08	96.00	137.00	159.00	179.50	246.00
Glucosa_en_Sangre	15.63	244.38	67.90	81.53	97.10	107.00	138.00
Hemoglobina_glicosilada	0.5979	0.3575	3.8000	5.0500	5.4000	5.9500	6.9000
Colesterol	54.07	2923.68	56.00	155.00	180.80	218.30	395.00
Presion_arterial_sistolica	23.31	543.18	90.00	120.00	125.00	140.00	237.00
Presion_arterial_diastolica	16.01	256.38	60.00	70.50	80.00	89.50	182.00
Trigliceridos	72.06	5192.56	28.00	106.55	152.00	202.55	381.00
Creatinina	0.931	0.867	0.500	0.795	0.900	1.100	6.600

Variable	Rango	IQR	Modo	N para moda	Asimetría	Curtosis
Edad	77.00	25.50	60	4	-0.12	-0.56
Peso	150.00	42.50	136; 152	3	0.41	0.21
Glucosa_en_Sangre	70.10	25.47	101	4	0.09	-0.44
Hemoglobina_glicosilada	3.1000	0.9000	5.4	7	0.12	-0.06
Colesterol	339.00	63.30	158	3	1.01	2.67
Presion_arterial_sistolica	147.00	20.00	120	15	1.82	5.89
Presion_arterial_diastolica	122.00	19.00	80	22	3.89	24.77
Trigliceridos	353.00	96.00	65; 71; 162.9	2	0.74	0.86
Creatinina	6.100	0.305	0.9	8	4.79	24.71

Fuente: Elaboración Propia.

Tabla 20 Estadística descriptiva asociada a pacientes con DM.

Variable	Conteo total	N	N*	NAcum	Porcentaje	PrcAcum	Media	Error estándar de la media
Edad	105	105	0	105	100	100	55.90	1.44
Peso	105	105	0	105	100	100	165.08	2.94
Glucosa en Sangre	105	105	0	105	100	100	210.0	11.5
Hemoglobina glicosilada	105	105	0	105	100	100	8.169	0.206
Colesterol	105	105	0	105	100	100	201.02	5.50
Presión arterial sistólica	105	105	0	105	100	100	137.06	2.43
Presión arterial	105	105	0	105	100	100	80.19	1.21

Variable	Conteo total	N	N*	NAcum	Porcentaje	PrcAcum	Media	Error estándar de la media
diastólica								
Triglicéridos	105	105	0	105	100	100	211.2	12.1
Creatinina	105	105	0	105	100	100	1.0628	0.0596

Variable	Desv.Est.	Varianza	Mínimo	Q1	Mediana	Q3	Máximo
Edad	14.72	216.76	15.00	47.00	56.00	65.50	89.00
Peso	30.14	908.29	109.35	144.00	160.00	186.00	263.00
Glucosa en Sangre	117.3	13769.1	66.2	139.5	174.0	258.9	1023.0
Hemoglobina glicosilada	2.111	4.454	5.500	6.825	7.400	9.050	15.400
Colesterol	56.38	3178.88	88.70	166.50	199.00	231.20	478.00
Presión arterial sistólica	24.86	618.02	100.00	120.00	130.00	150.00	230.00
Presión arterial diastólica	12.44	154.79	57.00	70.00	80.00	90.00	110.00
Triglicéridos	124.1	15398.3	51.0	138.5	188.3	233.8	840.0
Creatinina	0.6104	0.3725	0.2300	0.7550	1.0000	1.2000	4.9000

Variable	Rango	IQR	Moda	N para moda	Asimetría	Curtosis
Edad	74.00	18.50	56	7	-0.39	0.19
Peso	153.65	42.00	158	5	0.80	0.91
Glucosa en Sangre	956.8	119.4	166	3	3.52	21.31
Hemoglobina glicosilada	9.900	2.225	6.5	6	1.59	2.11
Colesterol	389.30	64.70	220; 228; 286	3	1.18	4.68
Presión arterial sistólica	130.00	30.00	120	21	1.40	2.41
Presión arterial diastólica	53.00	20.00	80	24	0.31	-0.53
Triglicéridos	789.0	95.3	145	3	2.63	9.36
Creatinina	4.6700	0.4450	1.3	10	4.34	24.69

Fuente: Elaboración Propia.

5.6.1 ANÁLISIS DE EDAD

El análisis del comportamiento de los datos en relación con la edad reveló que, de los 65 pacientes sin diagnóstico de DM, la media de edad fue de 54 años, lo que sugiere que esta es la edad promedio de los pacientes no diagnosticados con DM incluidos en la base de datos. Además, se observó una desviación estándar de 18.90, lo que indica que los datos no están muy concentrados cerca de la media (Ilustración 33).

El cálculo del error estándar de la media de edad dio como resultado 2.34, según se muestra en la Tabla 18. Este valor sugiere que existe variabilidad en los datos obtenidos en relación con la media, lo que indica que la estimación de la edad promedio de los pacientes sin DM no es precisamente exacta. Por otro lado, la varianza, al ser un valor elevado, confirma la gran dispersión de los datos y la diversidad de edades de los pacientes sin DM en la muestra.

Consecuente a lo anteriormente mencionado, se evidenció que el rango de edad es de 77 estableciendo al paciente más pequeño con 14 años y el mayor con 91 y que una de las edades se repitió 4 veces, siendo la moda 60. Por último, la curtosis fue negativa, de -0.56, lo que indicó que la distribución tiene una concentración más baja de valores alrededor de la media en comparación con una distribución normal estándar, y que las colas son menos pesadas que las de una distribución normal.

Los datos tuvieron una mayor diferencia en valores a la izquierda de la media, esto debido a que se obtuvo una asimetría de -0.12. En consecuencia, se observa que la cola de la izquierda es mayor a la cola de la derecha en el histograma (Ilustración 33).

Analizando las gráficas mostradas de los pacientes diabéticos en cuanto a la edad, se pudo concluir que, a diferencia de los pacientes no diagnosticados, la media fue de 55.9 con una desviación estándar de 14.72 y un error de 1.44 (Ilustración 35). Esto indica que los datos se encuentran mejor distribuidos que los pacientes no diagnosticados, observando que existieron agrupaciones más cercanas a la media (Ilustración 36).

La moda fue de 56 con una repetición de 7 veces junto con una curtosis de 0.19 en la distribución de edades. Exhibió que los datos tienen una distribución aproximadamente normal, es decir, la concentración de edades alrededor de la media es similar a la de una distribución

normal estándar. La distribución de edades tendría una forma de campana con colas no muy pesadas, lo que indica que hay una cantidad moderada de valores extremos. La asimetría fue de -0.39 indicando que los datos se encuentran recopilados a la izquierda de la campana, lo que indica que hay una concentración ligeramente mayor de valores más altos que de valores más bajos.

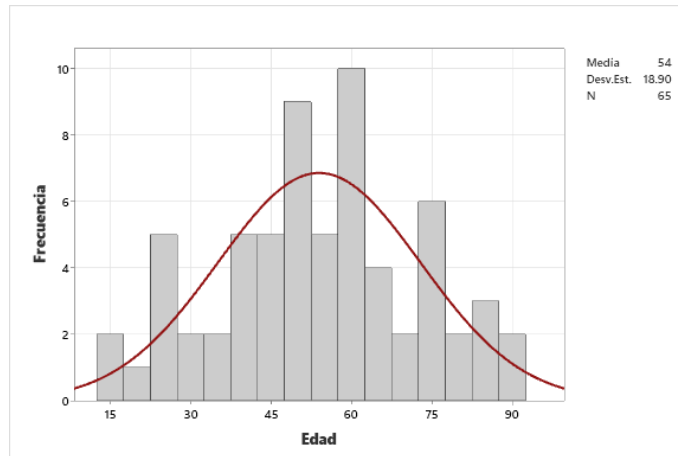


Ilustración 33 (Histograma con curva normal) de Edad asociado a pacientes sin diagnóstico de DM.

Fuente: Elaboración propia.

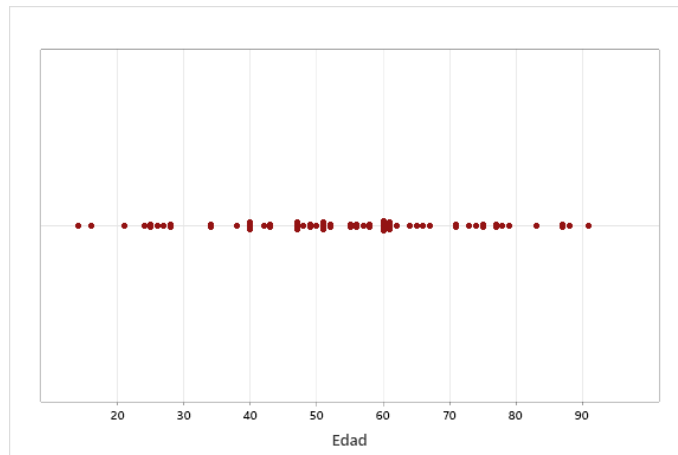


Ilustración 34 Gráfica de valores individuales de edad asociada a pacientes sin diagnóstico de DM.

Fuente: Elaboración propia.

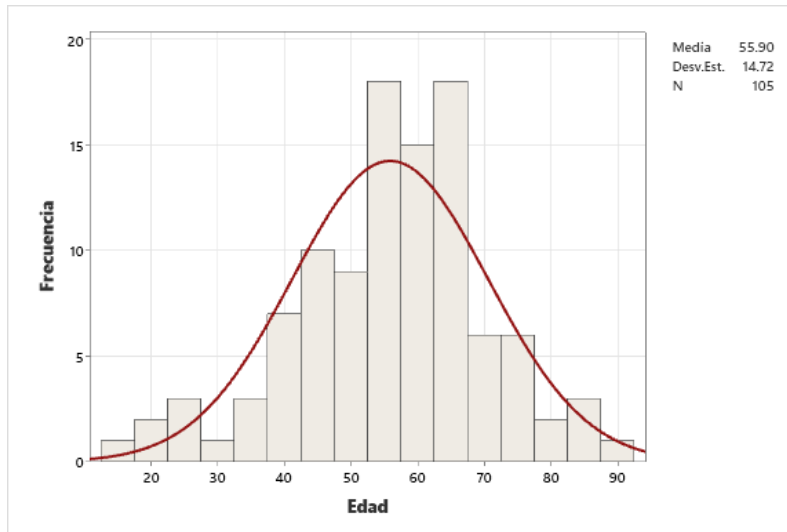


Ilustración 35 (Histograma con curva normal) de Edad asociado a pacientes con DM.

Fuente: Elaboración propia.

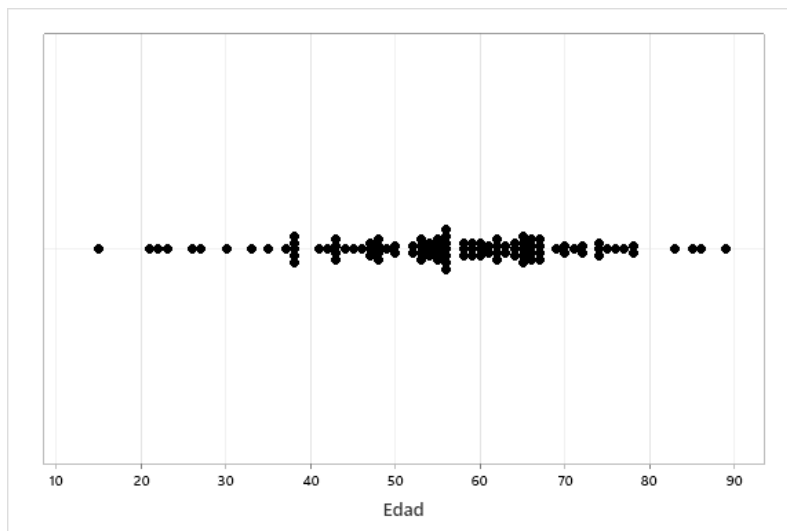


Ilustración 36 Gráfica de valores individuales de edad asociada a pacientes con DM.

Fuente: Elaboración propia.

5.6.2 ANÁLISIS DE PESO

Para el análisis del peso, se obtuvo una media de 160.9 libras, al igual que la edad se mostró una desviación estándar alta (31.61) lo que indica que los valores individuales están agrupados lejos de la media o dispersos (ilustración 37). El error estándar fue de 3.92, aún más grande que el error presentado en edad lo que indica que los datos de peso están todavía más alejados a la media y que la precisión es aún menor que en la edad, lo que hace que sea poco confiable (Ilustración 38).

De igual modo, se pudo observar en la tabla 13 que el valor mínimo de peso es de 96 y el mayor de 246, dando un rango de 150. La moda fue de 136 y 152 repitiéndose cada dato 3 veces. Con lo mencionado anteriormente es importante considerar que a pesar de que fueron datos aleatorios y que existe una variedad de pesos, la media que fue de 160.9 libras es un valor apropiado para pacientes sin un diagnóstico de diabetes, por lo que se pudo concluir que los datos a pesar de estar distribuidos dispersamente si fueron datos de peso y funcionales para la investigación.

La curtosis fue de 0.21 lo que al ser positivo significa que la distribución tiene una concentración de valores en la región central y colas más pesadas que la distribución normal (distribución leptocúrtica). En este caso, la asimetría fue de 0.41 indicando que la distribución está ligeramente inclinada hacia la derecha (asimetría positiva). En otras palabras, la cola derecha de la distribución es más larga que la cola izquierda, lo que indica que hay una concentración ligeramente mayor de valores más bajos que de valores más altos.

En el análisis de los datos de pacientes diagnosticados positivos a DM, se pudo concluir que la media fue de 165.08 con un error de 2.94 y una desviación estándar de 30.14, indicando que la distribución de los datos está muy alejada de la media (Ilustración 40). El peso mínimo estudiado fue de 109.35 y el máximo de 263. Es importante vigilar el peso de cada individuo pues, es muy común que los pacientes diabéticos también padezcan de otras enfermedades sistémicas como ser obesidad e hipertensión.

La gráfica de campana (Ilustración 39) mostró una asimetría de 0.80 y una curtosis de 0.91 (tabla 13) lo que indicó que los datos están recopilados más a la derecha de la campana,

significa que hay una concentración de valores más bajos en la cola izquierda de la distribución y una concentración de valores más altos en la cola derecha de la distribución. Por otro lado, la curtosis indicó que la distribución tiene una forma similar a una distribución normal, pero con colas ligeramente más pesadas. En otras palabras, la concentración de valores alrededor de la media es similar a la de una distribución normal, pero la distribución tiene una cantidad moderada de valores excesivos en ambos extremos.

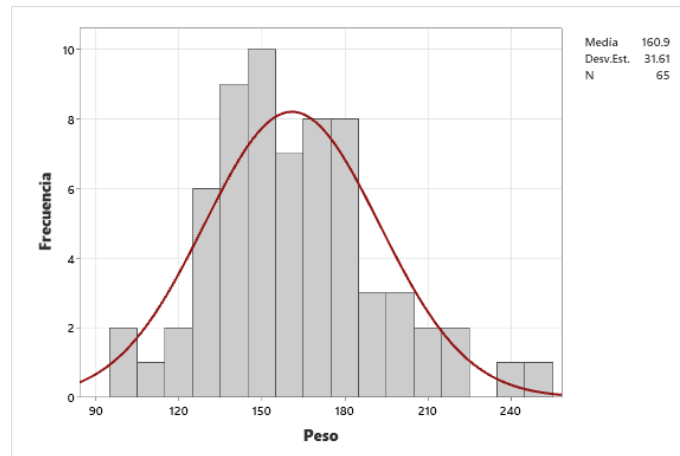


Ilustración 37 (Histograma con curva normal) de Peso asociado a pacientes sin diagnóstico de DM.

Fuente: Elaboración propia.

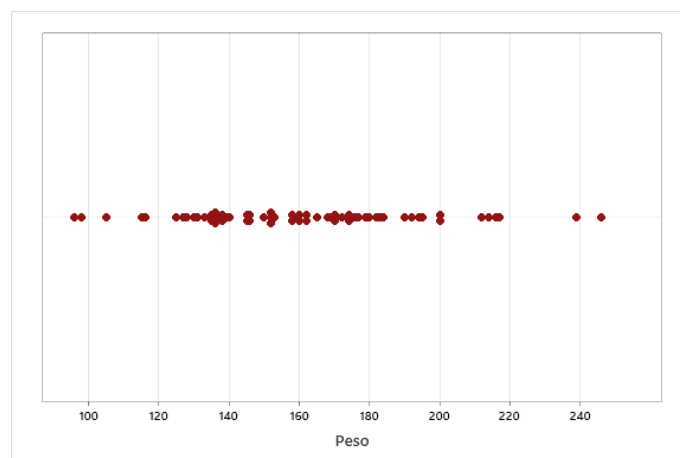


Ilustración 38 Gráfica de valores individuales de peso asociada a pacientes sin diagnóstico de DM.

Fuente: Elaboración Propia.

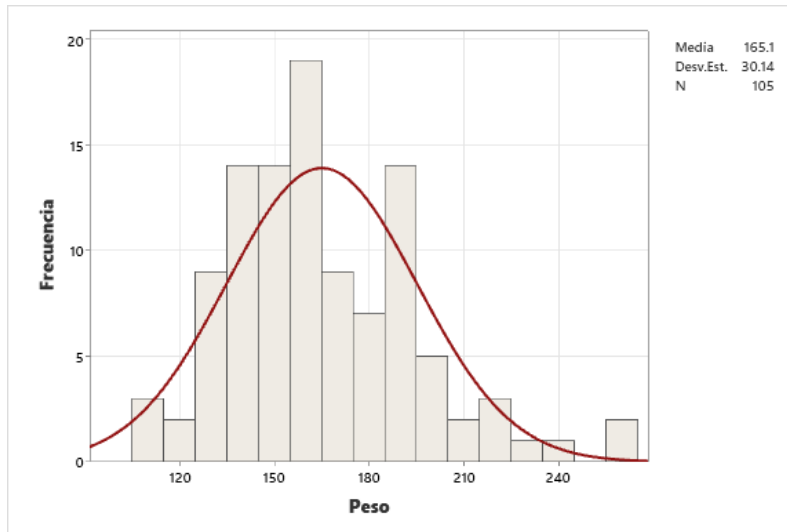


Ilustración 39 (Histograma con curva normal) de Peso asociado a pacientes con DM.

Fuente: Elaboración propia.

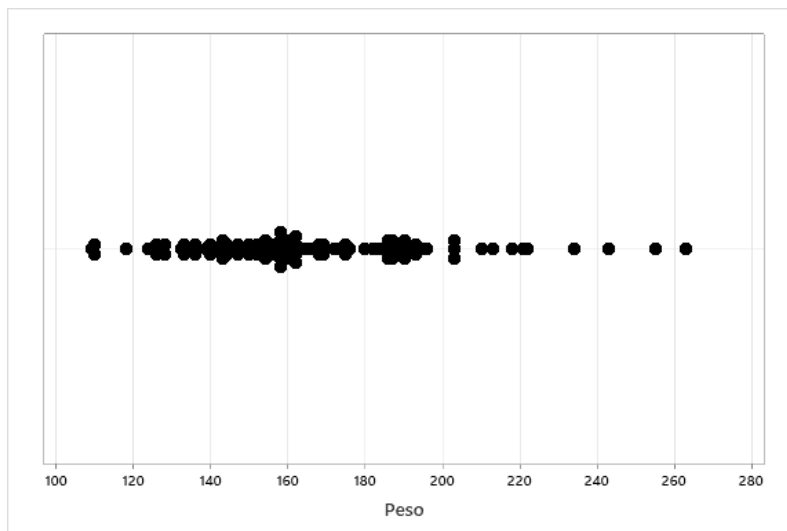


Ilustración 40 Gráfica de valores individuales de peso asociada a pacientes con DM.

Fuente: Elaboración Propia.

5.6.3 ANÁLISIS DE GLUCOSA EN SANGRE

La glucosa en sangre es una de las variables de más peso para el diagnóstico de la patología, pues un médico con solo observar sus valores y dirigido por unos rangos establecidos más el agregado de la hemoglobina glicosilada puede diagnosticar a un paciente, en este caso, la gráfica de campana (Ilustración 41) nos muestra que la media es de 95.89 que valida el hecho de que se analizaron datos de pacientes sin un diagnóstico positivo.

En esta ocasión la desviación estándar fue menor que en las variables anteriormente analizadas (15.63) lo que indica que los valores de glucosa se encuentran más cerca a la media en comparación a las anteriores (Ilustración 42).

El valor mínimo registrado fue de 67.90 y el máximo de 138 dando un rango de 70.10, de esto se pudo concluir que 138 es un dato atípico en esta muestra, pues es el valor frontera entre padecer y no padecer y 125, probablemente este sea un paciente que está controlado o que aún no ha sido diagnosticado, incluso podría ser prediabético, pero en vista de que la prediabetes no es un estado o una condición patológica se clasifica como no diabético.

La moda es 101, con una aparición de 4 veces y una curtosis de -0.44 (platicúrtica) lo que significa que la distribución tiene una forma más aplanada de lo que se esperaría en una distribución normal. Es decir, la distribución tiene una concentración moderada de valores en ambos extremos, lo que da como resultado colas más ligeras y una forma más aplanada.

La asimetría obtenida fue de 0.09 indicando que el histograma (Ilustración 41) tiene una cantidad similar de valores extremos a ambos lados de la media, lo que sugiere que la concentración de valores en la media es similar a la de una distribución normal.

En los pacientes diabéticos, la media obtenida fue de 210 con un error de 11.5 y una desviación de 117.3. Estos valores tan exagerados se deben a que hay valores atípicos dentro de la muestra, como ser un paciente un nivel de glucosa de 1023. En consecuencia, los datos se encuentran muy alejados de la media (Ilustración 44).

El mínimo fue de 66.2 y el máximo de 1023, estas fueron objeto de que un paciente diabético puede tener un nivel de glucosa bajo debido a un control, sin embargo, su nivel de hemoglobina glicosilada hace que se califique como diabético.

La moda fue de 166 con una aparición de 3 veces y una asimetría de 3.52 acompañada de una curtosis de 21.31 lo que nos indicó que la mayoría de los valores están concentrados en la cola izquierda de la distribución y hay una cantidad relativamente pequeña de valores extremos en la cola derecha y que la distribución es leptocúrtica (Ilustración 43).

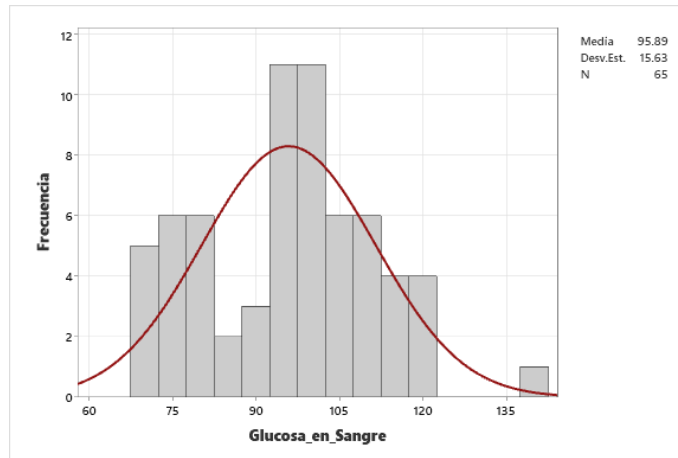


Ilustración 41 (Histograma con curva normal) de glucosa en sangre asociado a pacientes sin diagnóstico de DM.

Fuente: Elaboración Propia.

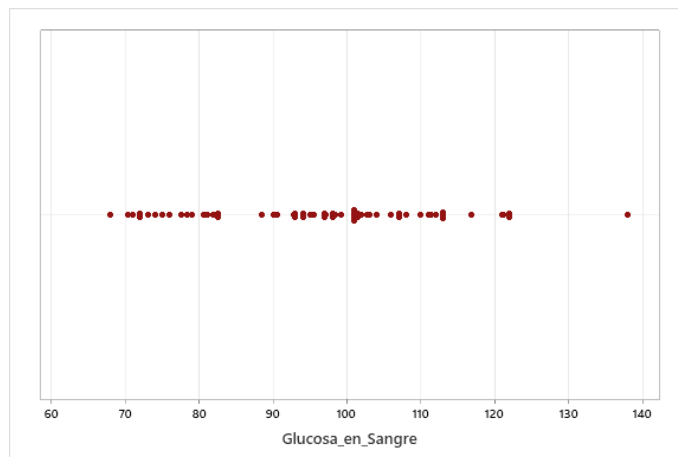


Ilustración 42 Gráfica de valores individuales de glucosa en sangre asociada a pacientes sin diagnóstico de DM

Fuente: Elaboración Propia.

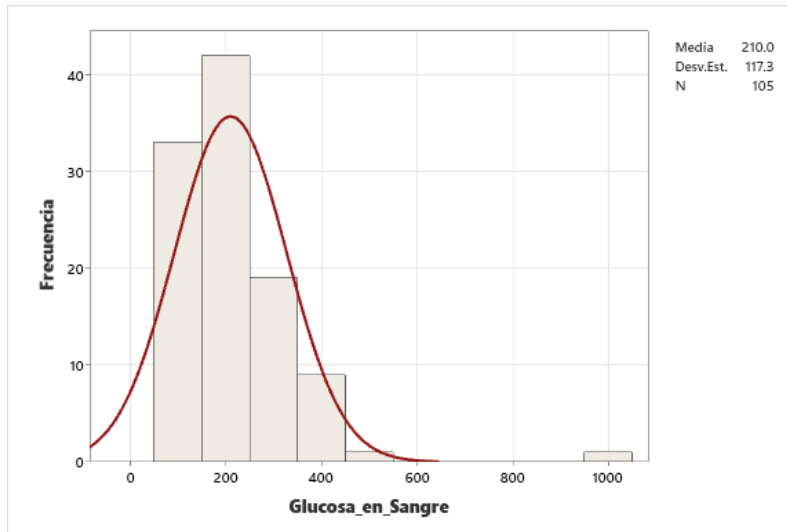


Ilustración 43 (Histograma con curva normal) de glucosa en sangre asociado a pacientes con DM.

Fuente: Elaboración Propia.

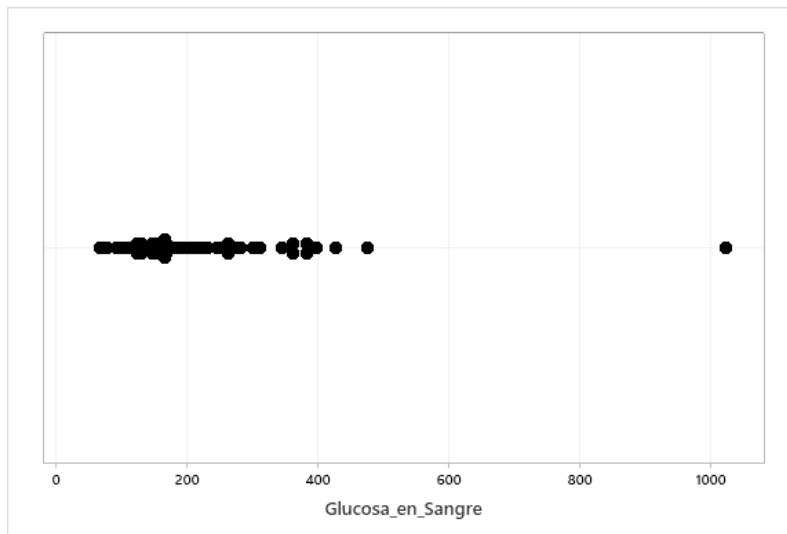


Ilustración 44 Gráfica de valores individuales de glucosa en sangre asociada a pacientes con DM

Fuente: Elaboración Propia.

5.6.4 ANÁLISIS DE HEMOGLOBINA GLICOSILADA

Al igual que la glucosa en sangre, la hemoglobina glicosilada juega un papel importante en el diagnóstico de la DM, en este caso, la media fue de 5.496 con una desviación estándar de 0.5979. Estos datos son buenos pues al obtener una desviación estándar menor, la fiabilidad y precisión de los datos aumentó y se evidenció que no se alejan mucho de la media (Ilustración 45 y 46).

El valor mínimo es de 3.8 y el máximo de 6.9, en esta variable el punto decisivo fue 6.5, pues al exceder este valor se considera que un paciente es diabético, en esta muestra solo hay pacientes sin el diagnóstico por lo que 6.9 es un dato atípico que influye en la desviación estándar. La moda fue de 5.4 con una aparición de 7 veces, este dato es coherente pues una persona sana aproximadamente ronda entre 5 y 6.4%. Por último, se obtuvo una curtosis de -0.06 que, aunque es casi 0, la curva de asimetría es negativa y posee una baja concentración en relación a la media aritmética (platicúrtica).

También se obtuvo una asimetría de 0.12 evidenciando que los datos están posicionados hacia la derecha de la campana (Ilustración 45) por lo que hay mayor número de valores en la cola derecha que en la cola izquierda. Los valores arriba de 6.5 indican que son pacientes diabéticos sin diagnóstico positivo.

Para los pacientes diabéticos se obtuvo una media de 8.169 con un error de 0.206 y una desviación de 2.11, reflejando que el comportamiento de los datos se encuentra dispersos alejados de la media debido a los datos atípicos encontrados (Ilustración 48). Se observó un mínimo de 5.50 y un máximo de 15.40 (valor atípico). La moda fue de 6.5, valor encontrado veces.

La asimetría fue de 1.59 y la curtosis de 2.11 lo que representa que los datos están ubicados a la izquierda de la campana (Ilustración 47) y que en la cola derecha hay una pequeña cantidad de valores, por otro lado, esta alta curtosis indicó una distribución muy leptocúrtica, lo que significa que la distribución tiene una forma puntiaguda y estrecha en comparación con una distribución normal.

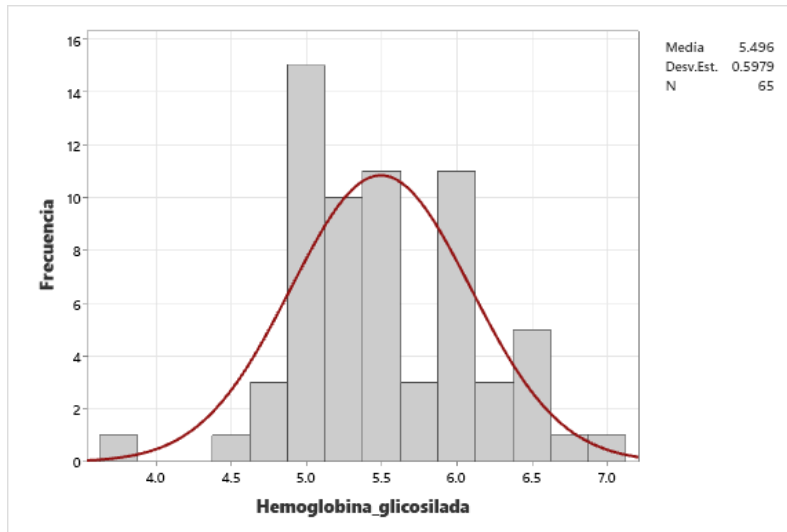


Ilustración 45 (Histograma con curva normal) de hemoglobina glicosilada asociado a pacientes sin diagnóstico de DM.

Fuente: Elaboración Propia.

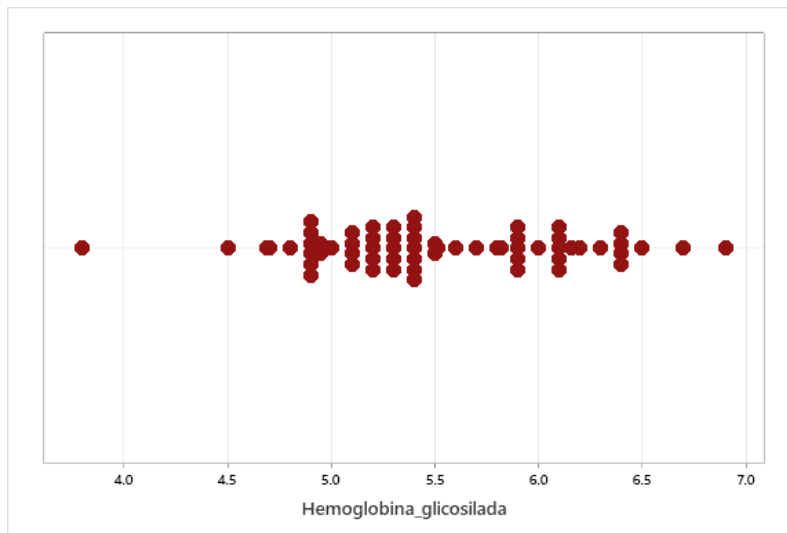


Ilustración 46 Gráfica de valores individuales de hemoglobina glicosilada asociada a pacientes sin diagnóstico de DM

Fuente: Elaboración Propia.

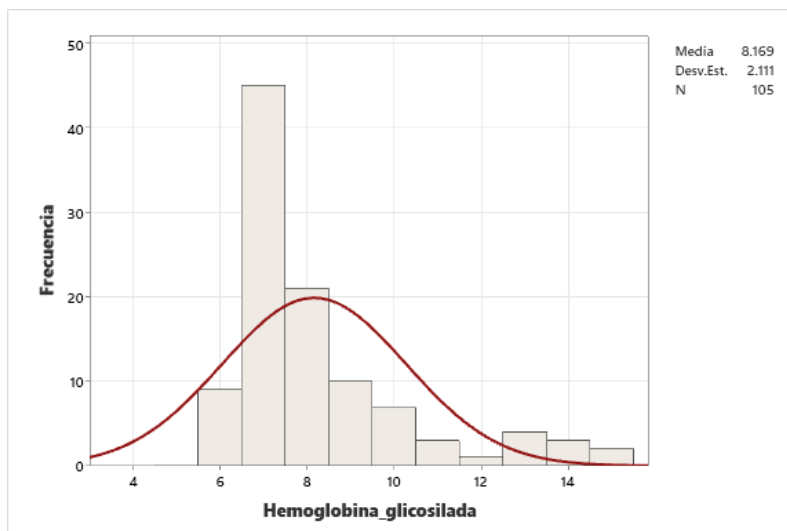


Ilustración 47 (Histograma con curva normal) de hemoglobina glicosilada asociado a pacientes con DM.

Fuente: Elaboración Propia.

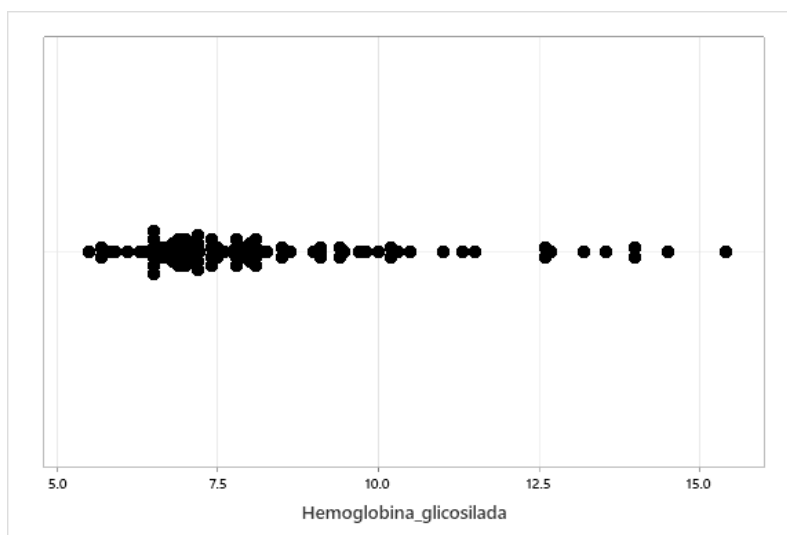


Ilustración 48 Gráfica de valores individuales de hemoglobina glicosilada asociada a pacientes sin con DM

Fuente: Elaboración Propia.

5.6.5 ANÁLISIS DE COLESTEROL

El histograma resultante de los análisis del colesterol expuso una media de 188.1 con una desviación estándar de 54.07, lo que indica que los datos se alejan bastante de la media, y esto se debe a que presenta valores atípicos de hasta 395 cuando clínicamente los valores aceptados son los menores a 200. Por eso, en el gráfico de valores individuales (Ilustración 50) se puede observar la distribución de los datos en un rango de 339.

La moda fue de 153 con repetición de 3 veces, la curva de campana presenta una curtosis de 2.67 lo que significa que tiene una curva de asimetría positiva y existe una gran concentración de valores en la región central de la distribución (Ilustración 49), lo que se denomina como leptocúrtica. La simetría fue de 1.1 mostrando que la mayoría de los datos se distribuyen en la cola izquierda de la campana y que hay una cantidad pequeña de datos en la cola derecha.

Para pacientes con diagnóstico positivo, la media del colesterol fue de 201.02 con un error de 5.5 y una desviación de 56.38 (51) evidenciando que al igual que en los pacientes no diabéticos los datos están un tanto alejados de la media y se pudieron observar algunos datos atípicos (Ilustración 52). Así mismo la varianza confirmó que los datos se encuentran muy dispersos y que existe una gran variabilidad de datos.

El valor mínimo analizado fue de 88.70 y el máximo de 478. Así mismo se obtuvo una moda de 220, 228 y 286 datos que se repitieron 3 veces cada uno. Estos valores son muy elevados debido a que lo normal es menor a 200.

La asimetría fue de 1.18 y la curtosis de 4.68 indicando que la mayoría de los datos se encuentran distribuidos en la cola izquierda de la campana lo que indica hubo más valores diferentes a la derecha de la media, evidenciando que los pacientes diabéticos suelen tener un colesterol elevado. La curtosis por otro lado indicó una distribución leptocúrtica, quiere decir que tiene una forma estrecha y puntiaguda en comparación con una distribución normal. El análisis de estos datos confirma la relación e influencia que tiene el colesterol en el diagnóstico de diabetes en la población estudiada.

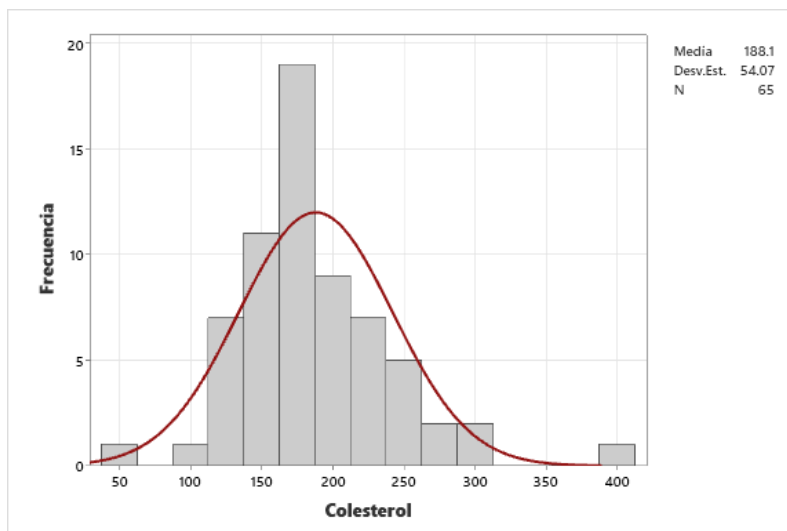


Ilustración 49 (Histograma con curva normal) de colesterol asociado a pacientes sin diagnóstico de DM.

Fuente: Elaboración Propia.

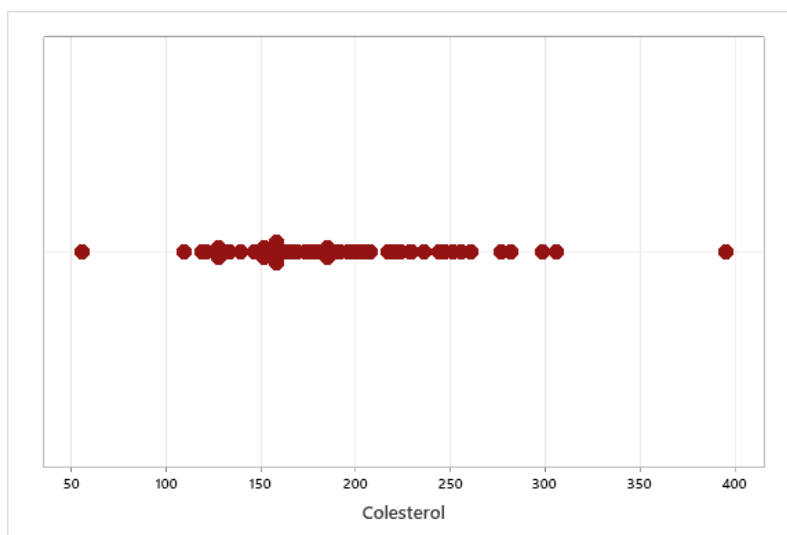


Ilustración 50 Gráfica de valores individuales de colesterol asociada a pacientes sin diagnóstico de DM

Fuente: Elaboración Propia.

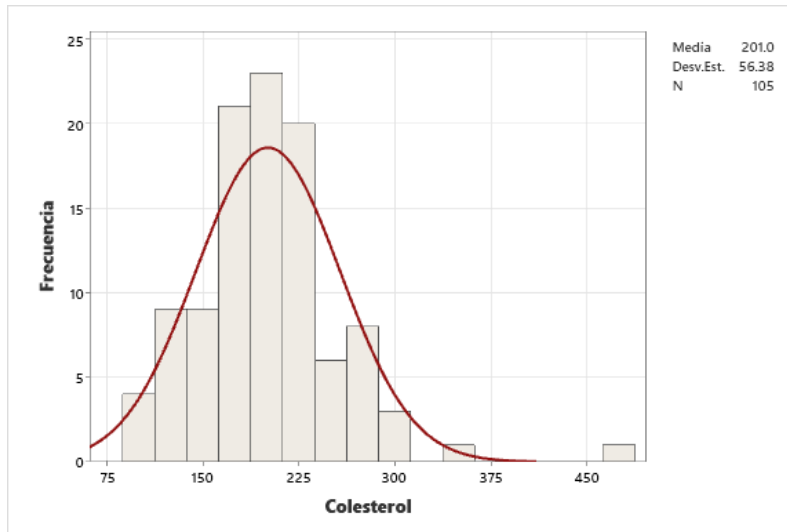


Ilustración 51 (Histograma con curva normal) de colesterol asociado a pacientes con DM.

Fuente: Elaboración Propia.

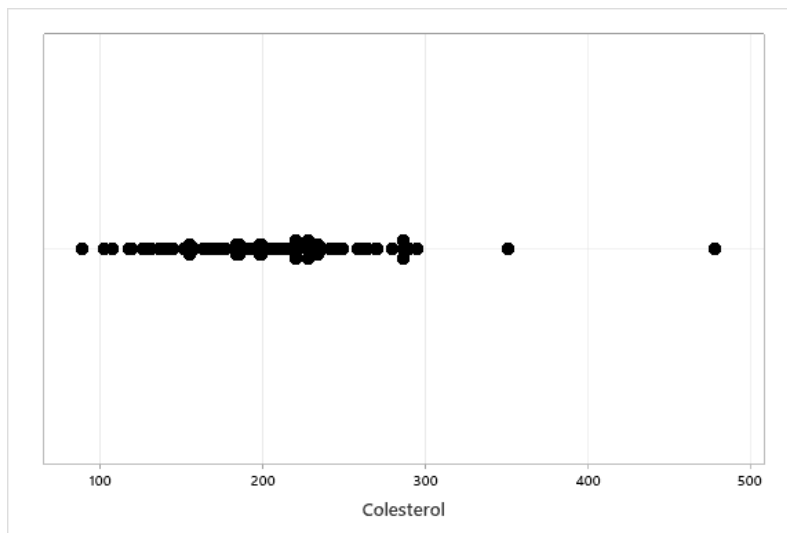


Ilustración 52 Gráfica de valores individuales de colesterol asociada a pacientes con DM

Fuente: Elaboración Propia.

5.6.6 ANÁLISIS DE PRESIÓN ARTERIAL SISTÓLICA

La presión arterial sistólica tuvo una media de 130.60 con una desviación estándar de 21.31 con un error del 2.89 según la tabla 12 y un 23.31 de desviación estándar, en el histograma (Ilustración 53) se pudo observar que existen valores atípicos muy elevados que ni siquiera se consideran dentro de la campana generada por el histograma, esto puede deberse a un error de medición o un caso especial. El valor ideal para medir la presión es 120/80 que serían 120 para presión arterial sistólica y 80 para presión arterial diastólica, según la tabla 12 el valor mínimo obtenido en la muestra fue de 90 y un máximo de 237, exponiendo un rango de 147. La presión arterial sistólica tuvo una media de 130.60 con una desviación estándar de 21.31.

La moda fue de 120, valor que se repitió 15 veces esto es normal ya que como se mencionó en personas sanas la presión debería manejarse en 120/80 o en un valor cercano, así mismo, en la gráfica de valores individuales se puede observar cómo en 120 están más agrupados los puntos (Ilustración 54).

La curtosis obtenida para esta variable fue de 5.89 indicando que tiene una curva asimétrica positiva por lo que la mayoría de los datos se observan en la parte izquierda del histograma y esto hace que sea leptocúrtica. Por otro lado, la asimetría fue de 1.82 lo que significa que los datos están sesgados a la derecha y que la mayor parte de los valores están concentrados en la cola izquierda de la campana (Ilustración 53).

Para los pacientes con diagnóstico positivo de DM se obtuvo una media de 137.06 con un error estándar de 2.44, una desviación de 24.86 y una varianza de 618.02 lo que indica que hay una gran cantidad de datos alejados de la media. El valor mínimo obtenido fue de 100 y el máximo fue de 230, este es un dato atípico y se puede observar que existen varios (Ilustración 56).

La moda fue de 120 repetido 21 veces, este es un indicativo positivo ya que 120 es el valor ideal para la PAS. La asimetría fue de 1.40 y la curtosis de 2.41, esto indica que hay muchos más valores a la derecha de la media y también mostró una distribución leptocúrtica (Ilustración 55).

Con esta información podemos concluir que los valores de presión arterial por encima de 130 pueden influir en el diagnóstico de DM, también evidencia que los pacientes con DM son propensos a padecer más enfermedades sistémicas, en este caso hipertensión pues, la PAS arriba de 140 el paciente es considerado hipertenso. Es normal observar un diagnóstico de DM más hipertensión arterial.

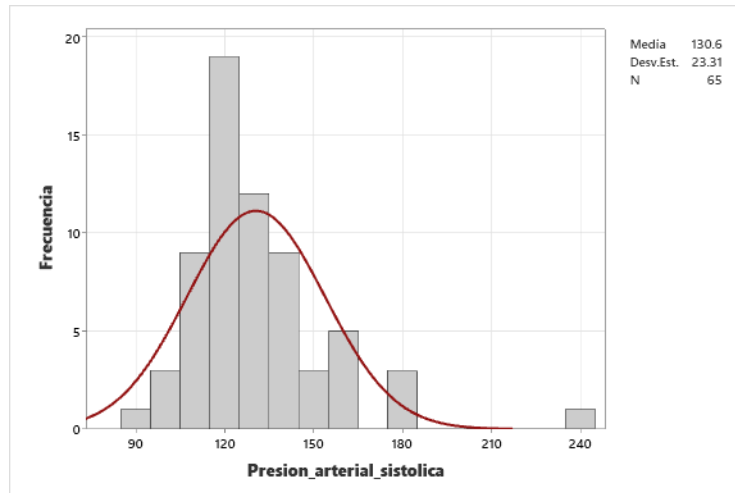


Ilustración 53 (Histograma con curva normal) de PAS asociado a pacientes sin diagnóstico de DM.

Fuente: Elaboración Propia.

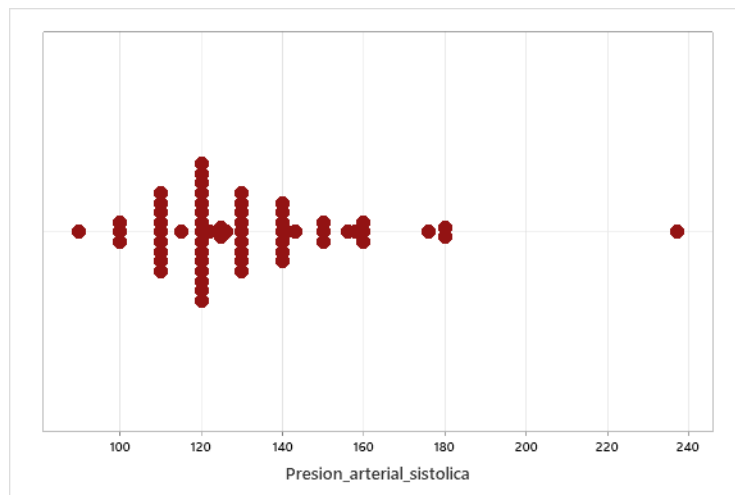


Ilustración 54 Gráfica de valores individuales de PAS asociada a pacientes sin diagnóstico de DM

Fuente: Elaboración Propia.

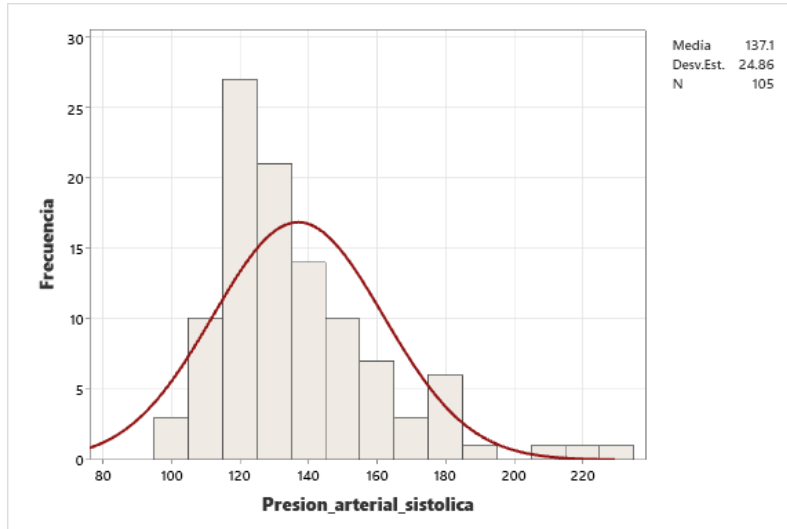


Ilustración 55 (Histograma con curva normal) de PAS asociado a pacientes con DM.

Fuente: Elaboración Propia.

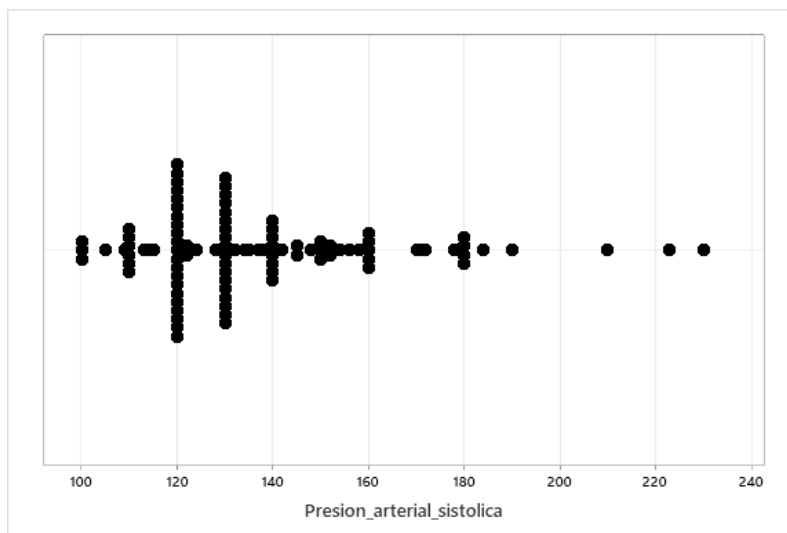


Ilustración 56 Gráfica de valores individuales de PAS asociada a pacientes con DM

Fuente: Elaboración Propia.

5.6.7 ANÁLISIS DE PRESIÓN ARTERIAL DIASTÓLICA

La media presentada para la variable de PAD fue de 80.75 con un error estándar de 1.99 y una desviación de 16.01, el comportamiento en comparación a la media lo podemos apreciar en el gráfico de campana (Ilustración 57) donde se pudo apreciar que la mayor cantidad de datos se encontró distribuida entre 60 y 100, demostrando que el valor mínimo es de 60 y el máximo es de 182, al igual que en la PAS se pueden apreciar valores atípicos que pueden alterar la fiabilidad de las variables en cuanto al análisis de cercanía a la muestra. Sin embargo, también es notorio que la moda fue de 80 en 22 personas (Ilustración 58), dato que es lo ideal para la PAD en una persona sana.

La curtosis de la PAD fue de 24.77 lo que indica que la distribución de los datos tiene una alta concentración de valores alrededor de la media y una gran cantidad de valores extremos en comparación con la distribución normal estándar. Esto sugiere que la distribución de los datos es muy leptocúrtica, lo que significa que tiene una concentración extrema de valores en el centro y colas muy pesadas. Añadiendo que la curva de asimetría es positiva por lo que los datos se ven agrupados a la izquierda del histograma, este tipo de distribución puede ser común en conjuntos de datos que tienen valores atípicos extremadamente grandes o pequeños, o que tienen una gran cantidad de valores extremos, como es el caso de PAD.

En pacientes con diagnóstico positivo la media obtenida fue de 80.19 con un error de 1.21 y una desviación estándar de 12.44 indicando que los datos se encuentran muy alejados de la media (Ilustración 60). Así mismo se pudo observar un mínimo de 57 y un máximo de 110. El valor ideal para la PAD es de 80 por lo que un valor arriba de 100 es bastante atípico, esto podría deberse a que el paciente es hipertenso o que la medición fue tomada de forma incorrecta.

La moda fue de 80 con una repetición de 24 veces. La asimetría mostrada fue de 0.31 y la curtosis de -0.53 lo que indica que en el histograma la cola de la derecha es más grande que la cola izquierda (Ilustración 59), sin embargo, no se observa tan pronunciada pues es una distribución casi normal. Así mismo, la curtosis nos indica que es una distribución platicúrtica lo

que significa que la distribución tiene una concentración moderada de valores extremos en ambos extremos, lo que da como resultado colas más ligeras y una forma más aplanada.

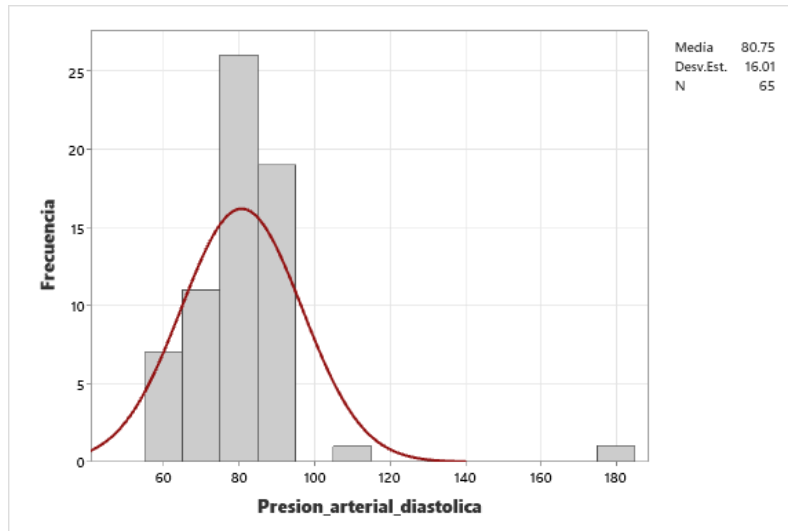


Ilustración 57 (Histograma con curva normal) de PAD asociado a pacientes sin diagnóstico de DM.

Fuente: Elaboración Propia.

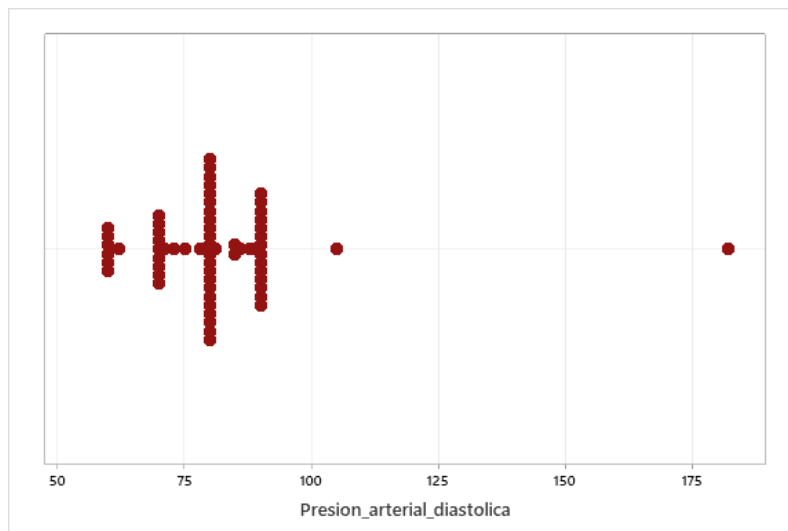


Ilustración 58 Gráfica de valores individuales de PAD asociada a pacientes sin diagnóstico de DM

Fuente: Elaboración Propia.

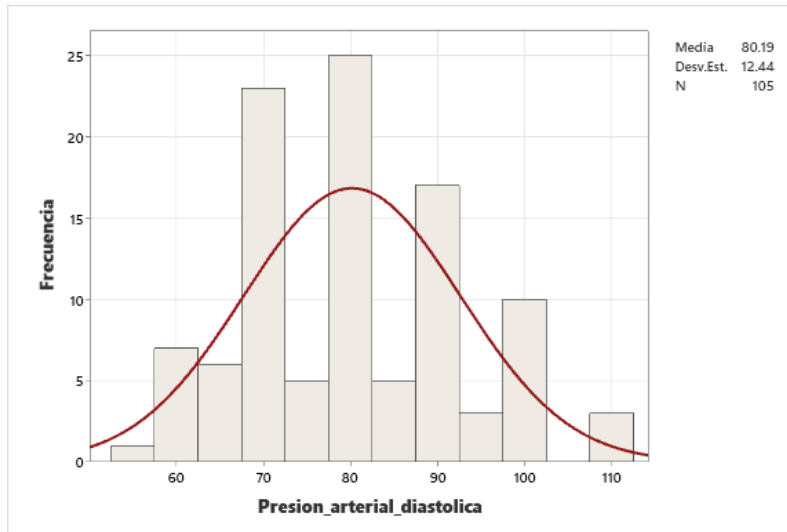


Ilustración 59 (Histograma con curva normal) de PAD asociado a pacientes con DM.

Fuente: Elaboración Propia.

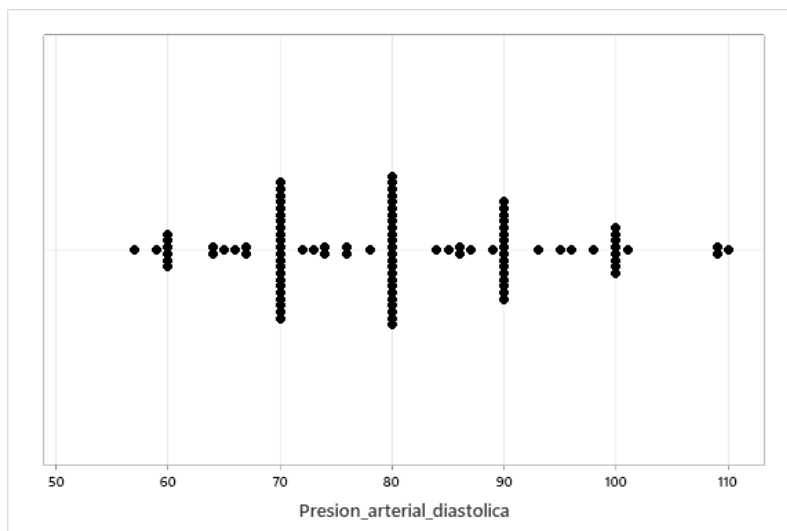


Ilustración 60 Gráfica de valores individuales de PAD asociada a pacientes con DM

Fuente: Elaboración Propia.

5.6.8 ANÁLISIS DE TRIGLICÉRIDOS

Para el análisis de triglicéridos, la tabla 12 muestra una media de 156.15 con un error de 8.94 y una desviación estándar de 72.06, esto quiere decir que los datos están muy alejados de la media tal como se observa en la gráfica de puntos (Ilustración 62) no hay datos acumulados sino que se observan como una línea recta. Así mismo la muestra presentó un valor mínimo de 28 y un máximo de 381, debido a la cantidad de valores atípicos es que se obtienen valores tan altos de error, desviación y rango que fue de 353.

Los datos que se repitieron fueron 65, 71 y 162.9 2 veces cada uno. Por último, la curtosis fue de 0.86 esto nos indica que la distribución de los datos tiene una concentración moderada alrededor de la media, con algunos valores extremos en las colas, pero no tan extremos como en una distribución con una curtosis alta. También tiene curva de asimetría positiva lo que indica que los datos están alineados a la izquierda y esto hace que sea leptocúrtica (Ilustración 61).

En el análisis de datos correspondiente a los pacientes con diagnóstico positivo de DM se pudo observar que la media fue de 211.2 con un error de 12.1, una desviación de 124.1 y una varianza de 15398.3 lo que indica que los datos están distribuidos muy alejados de la media (Ilustración 64) y que hay gran variabilidad de datos. El valor mínimo registrado fue de 51 y el máximo de 840, este valor es un dato muy atípico.

La moda fue de 145 con una aparición de 3 veces. La asimetría fue de 2.63 y la curtosis fue de 9.36. En consecuencia, los datos se encuentran mayor distribuidos en la cola izquierda de la distribución y hay una cantidad relativamente pequeña de valores extremos en la cola derecha. Así mismo la curtosis nos muestra que hay una distribución leptocúrtica lo que indica que la mayoría de las personas tiene valores similares (Ilustración 63).

Se puede concluir que el nivel de triglicéridos es influyente en el diagnóstico de DM pues el valor normal es menor igual a 150 y en este estudio se obtuvo una media de 211.2 lo que es considerado un valor alto.

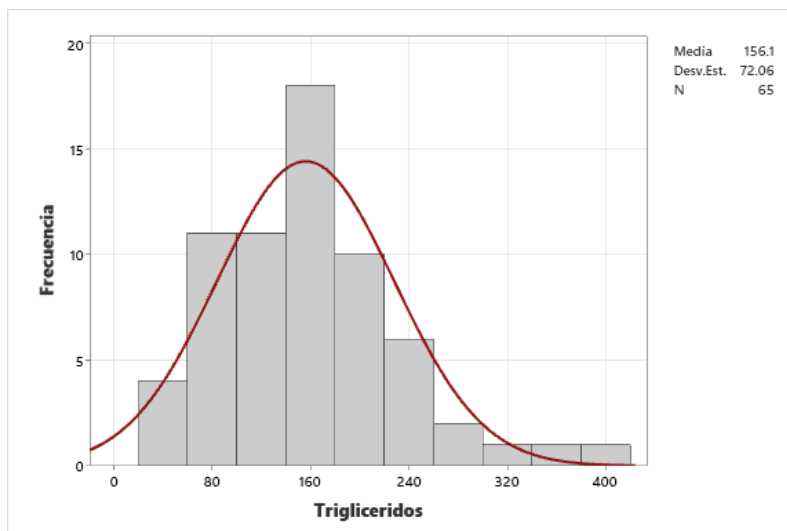


Ilustración 61 (Histograma con curva normal) de Triglicéridos asociado a pacientes sin diagnóstico de DM.

Fuente: Elaboración Propia.

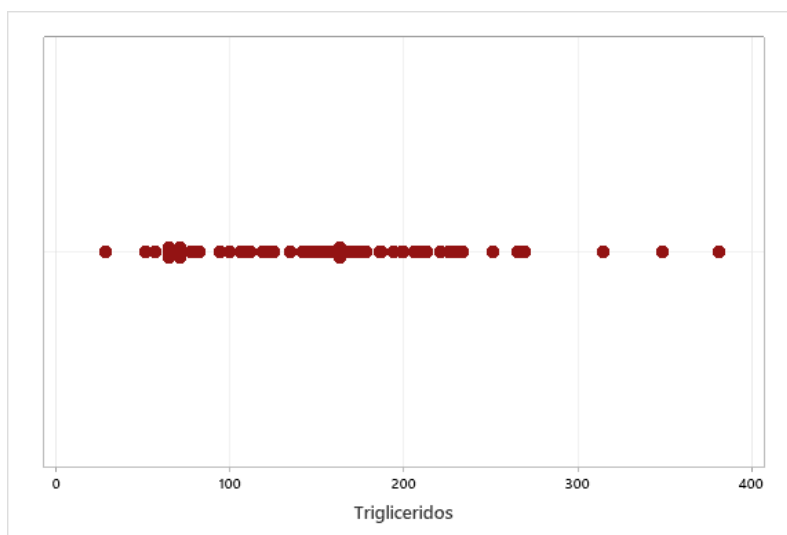


Ilustración 62 Gráfica de valores individuales de Triglicéridos asociada a pacientes sin diagnóstico de DM

Fuente: Elaboración Propia.

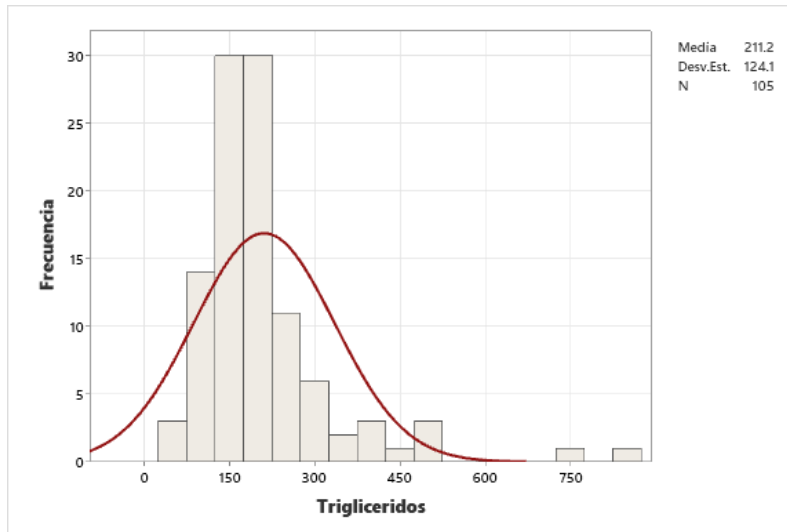


Ilustración 63 (Histograma con curva normal) de Triglicéridos asociado a pacientes con DM.

Fuente: Elaboración Propia.

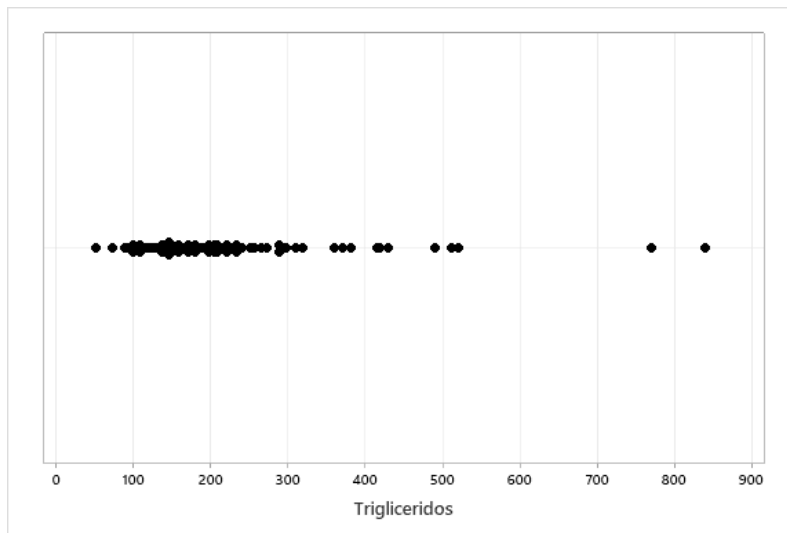


Ilustración 64 Gráfica de valores individuales de Triglicéridos asociada a pacientes con DM

Fuente: Elaboración Propia.

5.6.9 ANÁLISIS DE CREATININA

Los valores normales de creatinina se miden dependiendo del sexo, de 0.7 a 1.3 para hombres y 0.6 a 1.1 para mujeres, en este caso se obtuvo una media de 1.109 y ya que se desconoce la cantidad de hombres y mujeres es un dato funcional para ambos casos. La media se presentó con un error de 0.115 y una desviación de 0.931, al ser cifras pequeñas esto indica que los datos están cerca de la media tal y como se observa en la curva de campana (Ilustración 65) a excepción de algunos valores atípicos que se encuentran muy lejanos a la media, este tipo de valores puede pertenecer a pacientes que aún no hayan sido diagnosticados o a pacientes que están en el límite de padecer diabetes (Ilustración 66).

El rango obtenido fue de 6.10 con un valor máximo de 6.60 y un valor mínimo de 0.5, así mismo la moda fue de 0.9, dicho dato se repitió 8 veces. Por último, la curtosis fue de 24.71 lo que nos indica que la concentración es leptocúrtica lo que sugiere que los datos están concentrados en una distribución más alta y estrecha que una distribución normal. También se pudo observar que los datos están más a la izquierda lo que sugiere que tiene una asimetría positiva.

En contraste con lo anteriormente mencionado, en los pacientes con diagnóstico positivo de DM se obtuvo una media de 1.06 con un error de 0.059, una desviación estándar de 0.61 y una varianza de 0.372 lo que indica una distribución no muy alejada de la media a excepción de algunos valores atípicos (Ilustración 67). El valor mínimo registrado fue de 0.23 y el valor máximo fue de 4.9. La moda fue de 1.3 con una repetición de 10 veces (Ilustración 68).

La asimetría fue de 2.63 y la curtosis de 24.69 lo que indica que la distribución tiene más valores diferentes a la derecha de la media y una distribución leptocúrtica. Concluyendo se puede observar que la creatinina no tiene una influencia tan significativa pues la media no fue tan diferente en pacientes con diagnóstico positivo versus pacientes con diagnóstico negativo.

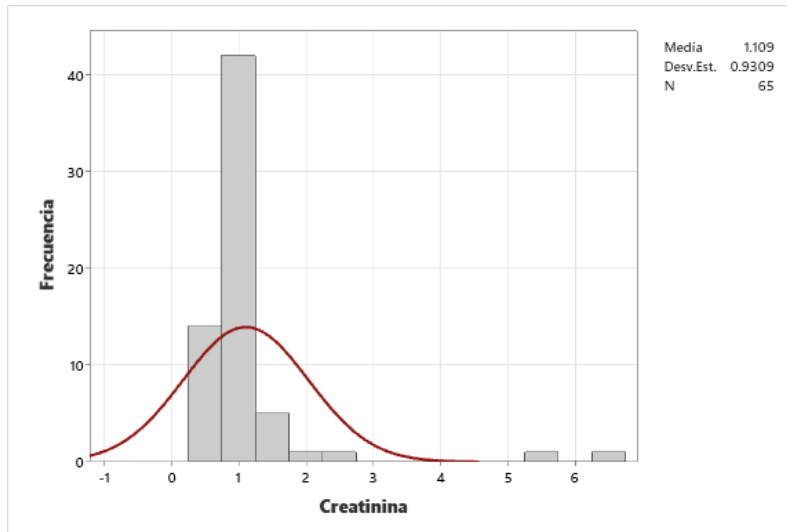


Ilustración 65 (Histograma con curva normal) de Creatinina asociado a pacientes sin diagnóstico de DM.

Fuente: Elaboración Propia.

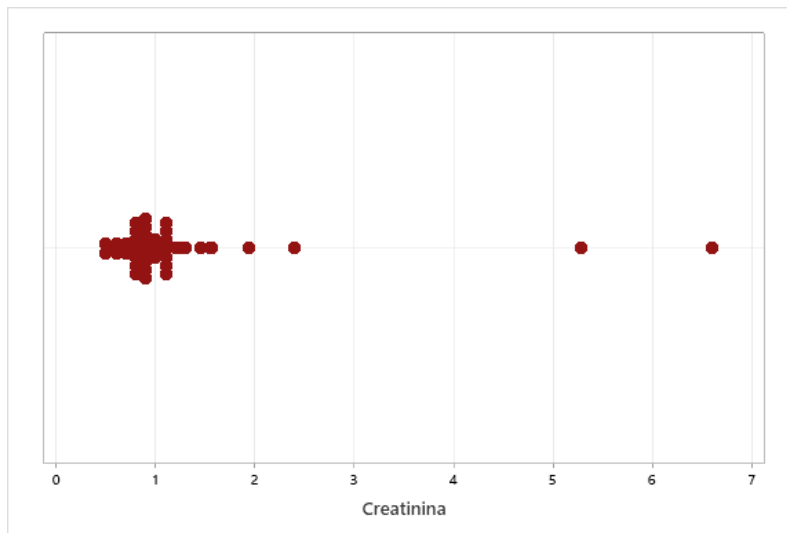


Ilustración 66 Gráfica de valores individuales de Creatinina asociada a pacientes sin diagnóstico de DM

Fuente: Elaboración Propia.

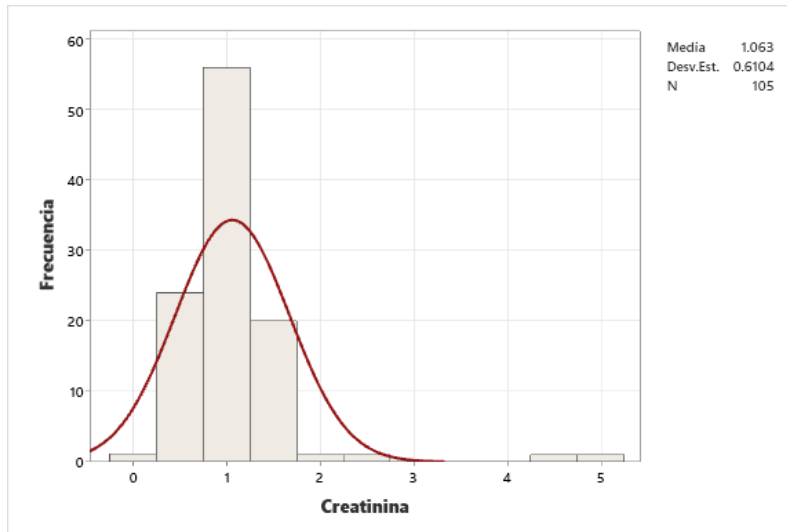


Ilustración 67 (Histograma con curva normal) de Creatinina asociado a pacientes con DM.

Fuente: Elaboración Propia.

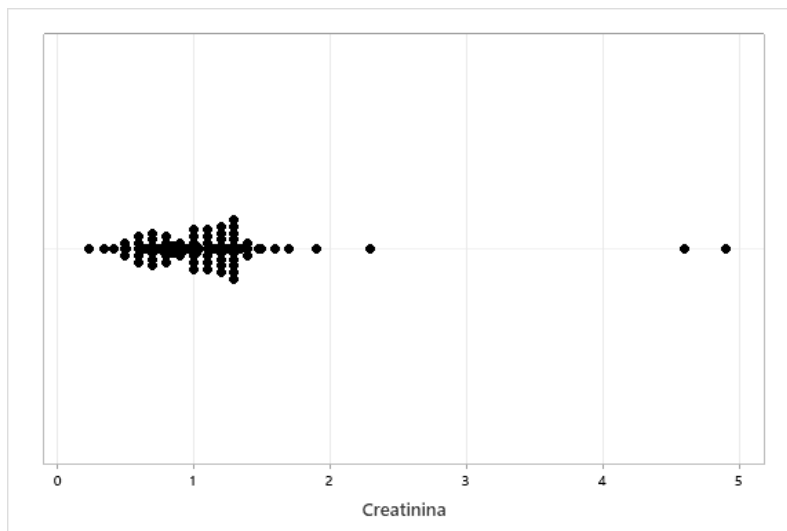


Ilustración 68 Gráfica de valores individuales de Creatinina asociada a pacientes con DM

Fuente: Elaboración Propia.

VI. CONCLUSIONES

1. Según los resultados obtenidos, se puede inferir que es factible y eficaz elaborar un algoritmo computacional como herramienta de diagnóstico para la Diabetes Mellitus con capacidades clínicas probabilísticas. La capacidad de los algoritmos para procesar grandes cantidades de datos y utilizar técnicas probabilísticas permite aumentar la precisión en el diagnóstico de la diabetes, lo que puede impactar positivamente en la calidad de vida de los pacientes orientando al diagnóstico temprano y prevención de la enfermedad. Además, la combinación de la inteligencia artificial y la medicina puede abrir nuevas posibilidades para el diagnóstico y tratamiento de una amplia gama de enfermedades crónicas, agilizar procesos y apoyar a médicos en la toma de decisiones.
2. El uso de IA como método diagnóstico es un tema de investigación en constante evolución. En general, los expertos están de acuerdo en que la IA tiene el potencial de mejorar la precisión y eficiencia del diagnóstico en muchos campos de la medicina. Sin embargo, se puede concluir que la mayoría de expertos en el tema no consideran el uso de aprendizaje automático para brindar un diagnóstico probabilístico, sino que utilizan un enfoque aplicado a la clasificación del tipo de Diabetes (Tipo I y Tipo II) y la predicción de las posibles complicaciones.
3. La implementación de la inteligencia artificial como método diagnóstico y probabilístico de la Diabetes Mellitus ha demostrado resultados prometedores. Aunque existen limitaciones y desafíos en la implementación de algoritmos de aprendizaje automático para proporcionar diagnósticos probabilísticos precisos, se ha comprobado que ciertas técnicas de inteligencia artificial pueden resultar más efectivas que otras. En cuanto a los tres métodos estudiados, se evidenció mayor precisión para la clasificación el método de árbol de decisión con la limitante de que no proporciona un valor probabilístico. Y, a pesar de que la red neuronal artificial tuvo un menor desempeño en ese aspecto, tiene mayor utilidad debido al porcentaje probabilístico.
4. Al comprender y aplicar los principios computacionales y matemáticos detrás de una red neuronal, los profesionales en el campo del análisis de datos pueden crear modelos precisos y eficaces para realizar clasificaciones y predecir enfermedades. Los principios

matemáticos y computacionales más relevantes involucrados en el proceso de desarrollo de una red neuronal artificial se fundamentan principalmente en la etapa del entrenamiento, donde se busca mejorar la precisión del modelo. En esta, a través de un algoritmo computacional se realiza el paso de información de una neurona a otra por medio de una función matemática, conocida como función de activación, que establece un valor que debe ser superado para permitir la transmisión de información. De igual manera, utilizando la derivada de la función de activación se propaga el error hacia atrás entre las capas de la red para así lograr disminuirlo.

5. Se desarrolló la RNA bajo el funcionamiento lógico de las librerías TensorFlow y Keras, al finalizar se determinó que el uso de las mismas puede tener un efecto significativo en la eficiencia y precisión del modelo. Una de las principales ventajas identificadas al trabajar con librerías preexistentes es que permite realizar tareas complejas de manera más rápida y sencilla, facilitando el proceso de desarrollo de la red para brindar un diagnóstico preciso de Diabetes Mellitus e identificar la probabilidad de padecer la patología. Esto tiene un impacto significativo especialmente cuando se trata de trabajos de corta duración.

VII. RECOMENDACIONES

1. Se sugiere la expansión de la base de datos empleada para entrenar la red neuronal, con la finalidad de mejorar sus capacidades diagnósticas y probabilísticas. Esto permitiría aumentar la precisión y confiabilidad de los resultados obtenidos por la red neuronal, a través de la inclusión de datos adicionales que reflejen la variabilidad y complejidad de la enfermedad. La ampliación de la base de datos puede lograrse mediante la inclusión de datos de pacientes de diversas poblaciones, así como la incorporación de nuevos factores de riesgo y comorbilidades asociadas con la diabetes. El resultado final sería una red neuronal más precisa y eficiente, que podría ser de gran utilidad en la detección temprana y tratamiento de la diabetes.
2. Se recomienda encarecidamente que, independientemente del resultado obtenido por la red neuronal en la interfaz gráfica desarrollada, se busque la atención de un centro de salud integral para confirmar el diagnóstico y establecer un tratamiento adecuado para la situación de salud del paciente. Es importante tener en cuenta que la red neuronal es una herramienta complementaria para la toma de decisiones médicas, y que no debe reemplazar el juicio clínico y la experiencia del profesional de la salud. Por lo tanto, se debe buscar siempre la evaluación médica y el seguimiento apropiado para asegurar el mejor cuidado posible del paciente.
3. Se sugiere desarrollar una aplicación de interfaz portátil para que los usuarios tengan acceso a ella de manera libre y sin costo alguno.
4. Se recomienda extender la investigación sobre variables asociadas a la patología, tales como el índice de masa corporal y los hábitos relacionados con la actividad física.

VIII. APLICABILIDAD/ IMPLEMENTACIÓN

El prototipo de red neuronal artificial puede ser utilizado principalmente por aquella población en riesgo de padecer la patología con el fin de realizar cierto tamizaje previo a recibir atención médica por parte de un profesional experto. En esta población en riesgo se encuentran familiares de pacientes diabéticos, personas hipertensas o cualquier persona que presente indicios o sospechas de padecer la enfermedad. De esta manera, se busca facilitar la identificación temprana de la diabetes, lo que podría contribuir a un diagnóstico y tratamiento oportunos. Asimismo, el prototipo puede ayudar a crear conciencia sobre los síntomas y factores de riesgo asociados con la diabetes, lo que podría llevar a una mejor prevención y gestión de esta enfermedad crónica.

IX. EVOLUCIÓN DE TRABAJO ACTUAL / TRABAJO FUTURO

1. Se plantea la viabilidad de incorporar la telemedicina en la atención médica, utilizando un registro médico electrónico que contenga la información que ha sido proporcionada a la red neuronal. Esto permitiría al médico verificar el diagnóstico otorgado por la red, y, además, la telemedicina permite una interacción directa entre el médico y el paciente, lo que hace posible la provisión de recomendaciones y tratamientos sin que el paciente tenga que visitar el establecimiento de salud físicamente.
2. Dado que la diabetes es una enfermedad sistémica que se relaciona directa e indirectamente con la hipertensión y la obesidad, se propone la posibilidad de entrenar un algoritmo con datos de estas patologías para identificar las interconexiones existentes entre ellas. Esto permitiría al algoritmo diagnosticar y brindar una probabilidad de padecer las tres patologías, ofreciendo un enfoque integrado y holístico para el tratamiento de estas afecciones. La combinación de datos de múltiples fuentes y la aplicación de técnicas de aprendizaje automático permitiría descubrir patrones y relaciones que podrían ser útiles para mejorar la comprensión de estas patologías y ofrecer una atención más eficaz y personalizada a los pacientes afectados.

X. CONCORDANCIA DE LOS SEGMENTOS DE LA TESIS CON LA INVESTIGACIÓN

La matriz de concordancia de los segmentos de la tesis con la investigación expone brevemente lo desarrollado en cada capítulo de la investigación (Tabla 21). Esta permite tener un panorama más claro y visualizar fácilmente la concordancia que existe entre ellos.

Tabla 21 Concordancia de los segmentos de la tesis con la investigación.

Capítulo I		Capítulo II		Capítulo III
Título de Investigación	Introducción	Planteamiento del Problema		Estado del Arte
Desarrollo de una red neuronal artificial para el diagnóstico clínico y probabilístico de Diabetes Mellitus.	Actualmente, las redes neuronales han generado un gran impacto en distintos campos, incluyendo la medicina. La diabetes mellitus es una enfermedad que afecta a la gran mayoría de la población a nivel mundial. Aunque en la actualidad ya existen diversos aparatos utilizados para llevar un mejor control de la enfermedad, la tendencia es que el número de casos continúe en aumento.	Problema de Investigación	Pregunta General	Objetivo General
		Aunque la glucemia plasmática tiene un gran impacto a la hora de diagnosticar la DM, con ayuda de la inteligencia artificial se puede brindar un apoyo al personal de salud agilizando los procesos con el mínimo porcentaje de error. Por otro lado, la glucemia capilar es el método diagnóstico más sobresaliente debido a la rapidez en la obtención del resultado, sin	¿Cómo desarrollar un método de diagnóstico para la Diabetes Mellitus con capacidades clínicas probabilísticas utilizando inteligencia artificial?	Desarrollar un método de diagnóstico para la Diabetes Mellitus con capacidades clínicas probabilísticas utilizando inteligencia artificial a través de un algoritmo computacional desarrollado en lenguaje de programación Python.
		Preguntas Específicas	Objetivos Específicos	
		¿Cuál es el enfoque de aplicabilidad que se le ha dado al aprendizaje automático en el contexto clínico diagnóstico de la Diabetes Mellitus por parte de otros	Recopilar información sobre aplicación del aprendizaje automático por parte de otros investigadores en el contexto clínico diagnóstico de la Diabetes Mellitus desde el año 2013 hasta la actualidad.	En el estado del arte se realiza una análisis de la situación actual de la Diabetes Mellitus tomando dos panoramas a analizar, el microentorno y macroentorno. En este se determinó la prevalencia de la patología y costos asociados a la misma a nivel mundial, y luego, en Centroamérica específicamente Honduras. De igual manera, se abordaron conceptos relacionados a la patología, la

<p>El proyecto de investigación se realizará con el fin de desarrollar una red neuronal que permita el diagnóstico y probabilidad de padecer diabetes de forma rápida y precisa.</p>	<p>embargo, este también tiene sus limitaciones que pueden afectar la validez de la lectura.</p>	<p>investigadores a partir del año 2013?</p>	<p>clasificación de la misma y los métodos diagnósticos actuales. Finalmente se describen los fundamentos de la inteligencia artificial y algunos modelos que pueden ser aplicados para realizar un diagnóstico de Diabetes Mellitus, incluyendo sus principios básicos.</p>
		<p>¿Cuál es la funcionalidad y aplicabilidad de tres distintos métodos de inteligencia artificial como herramientas diagnósticas y probabilísticas de la Diabetes Mellitus?</p>	<p>Evaluar la funcionalidad y aplicabilidad de tres métodos de inteligencia artificial como herramientas diagnósticas y probabilísticas de la Diabetes Mellitus.</p>
		<p>¿Cuáles son los principios computacionales y matemáticos que se pueden aplicar en el desarrollo de un red neuronal?</p>	<p>Identificar los principios computacionales y matemáticos que se pueden aplicar en el desarrollo de una red neuronal alimentada con una muestra de datos de pacientes hondureños.</p>
		<p>¿Qué efecto tiene el uso de librerías durante el desarrollo de una red neuronal para diagnosticar la Diabetes Mellitus y determinar la probabilidad de padecer la patología?</p>	<p>Determinar el efecto que tiene el uso de librerías durante el desarrollo de una red neuronal para diagnosticar la Diabetes Mellitus y determinar la probabilidad de padecer la patología.</p>

	Capítulo IV	Capítulo V	Capítulo VI	Capítulo VII
Variables de Investigación	Metodología de Estudio	Resultados y Análisis	Conclusiones	Recomendaciones
Dependiente Red neuronal artificial	Se utilizó la metodología en v por la creación del prototipo de software, en este caso una red neuronal para el diagnóstico clínico y probabilístico de Diabetes Mellitus.	Tras haber evaluado tres modelos de aprendizaje automático, se determinó que el árbol de decisión es más eficiente para realizar la clasificación de diagnóstico (positivo y negativo). Sin embargo, carece de la capacidad de brindar un porcentaje probabilístico de padecer la patología a diferencia de la red neuronal.	La IA como método diagnóstico resultó ser una herramienta bastante precisa para la detección de Diabetes Mellitus con capacidades clínicas probabilísticas. A pesar de no ser utilizada con este mismo enfoque por otros investigadores, si se ha visto cada vez más la aplicación de la inteligencia artificial en el ámbito de la salud. Sobre todo, únicamente con un propósito de clasificación. Para la creación de un prototipo de red neuronal que proporcione un diagnóstico clínico y probabilístico de DM se debe considerar que en este interactúan un conjunto de principios matemáticos y computacionales esenciales para el funcionamiento adecuado de la misma.	Realizar una expansión de la base de datos empleada para entrenar la red neuronal, con la finalidad de mejorar las capacidades diagnósticas y probabilísticas de la misma.
Independientes Variables de entrada de la red. Modelos de aprendizaje automático. Arquitectura de la red. Datos de entrenamiento.	En esta, la fase de diseño se basó en la identificación de variables a utilizar en la red neuronal, la selección del lenguaje de programación e IDE y la definición de la arquitectura de la red. En la fase de validación se realizó el entrenamiento de los métodos de aprendizaje automático para validar la coherencia de los resultados brindados por la red comparando con el método de mayor precisión. También se realizó un análisis de los valores predictivos y diagnósticos y, finalmente, se desarrolló la interfaz gráfica.	La red neuronal desarrollada con las librerías TensorFlow y Keras obtuvo un porcentaje de precisión de 91% para datos nuevos y 94% para datos conocidos. Por lo que se puede decir que los resultados son bastante confiables.	Además, se recomienda buscar atención médica en un centro de salud integral para confirmar el diagnóstico y establecer un tratamiento adecuado para la situación de salud del paciente.	

Capítulo VIII	Capítulo IX
Aplicabilidad/ Implementación	Evolución de trabajo actual/ Trabajo a futuro
<p>Se determinó que el prototipo de red neuronal artificial puede ser utilizado por aquella población en riesgo de padecer la patología con el fin de realizar cierto tamizaje o actuar como un triaje previo a recibir atención médica por parte de un profesional experto. Como ser familiares de pacientes diabéticos, personas hipertensas o cualquier persona que presente indicios o sospechas de padecer la enfermedad.</p>	<p>Se plantea la viabilidad de incorporar al prototipo la telemedicina, utilizando un registro médico electrónico que contenga la información que ha sido proporcionada a la red neuronal permitiéndole al médico verificar el diagnóstico otorgado por la red y una interacción directa con el paciente para brindarle observaciones y/o tratamiento.</p> <p>Además, se plantea la posibilidad de incorporar un diagnóstico clínico y probabilístico de enfermedades que están estrechamente relacionadas con la DM, como ser la hipertensión y la obesidad.</p>

Fuente: Elaboración propia.

BIBLIOGRAFÍA

1. Agudelo-Botero, M., & Dávila-Cervantes, C. A. (2015). Carga de la mortalidad por diabetes mellitus en América Latina 2000-2011: Los casos de Argentina, Chile, Colombia y México. *Gaceta Sanitaria*, 29(3), 172-177. <https://doi.org/10.1016/j.gaceta.2015.01.015>
2. Alam, U., Asghar, O., Azmi, S., & Malik, R. A. (2014). Chapter 15—General aspects of diabetes mellitus. En D. W. Zochodne & R. A. Malik (Eds.), *Handbook of Clinical Neurology* (Vol. 126, pp. 211-222). Elsevier. <https://doi.org/10.1016/B978-0-444-53480-4.00015-1>
3. American Diabetes Association. (2022). Standards of Medical Care in Diabetes—2022 Abridged for Primary Care Providers. *Clinical Diabetes*, 40(1), 10-38. <https://doi.org/10.2337/cd22-as01>
4. ANSI/CTA Standard the Use of Artificial Intelligence in Health Care: Trustworthiness, n.º ANSI/CTA-2090 (2021). https://www.imagwiki.nibib.nih.gov/sites/default/files/Standards_ANSI-CTA-2090_0.pdf
5. Aracena, C., Villena, F., Arias, F., & Dunstan, J. (2022, octubre 26). *Aplicaciones de aprendizaje automático en salud | Elsevier Enhanced Reader*. REVISTA MÉDICA CLÍNICA LAS CONDES. <https://doi.org/10.1016/j.rmcl.2022.10.001>
6. ASALE, R.-, & RAE. (s. f.). *Diccionario de la lengua española*. «Diccionario de la lengua española» - Edición del Tricentenario. Recuperado 10 de mayo de 2023, de <https://dle.rae.es/hiperglucemia>
7. Atención a los pacientes diabéticos y a personas con problemas de salud mental en tiempos de COVID-19 en Honduras—OPS/OMS | Organización Panamericana de la Salud. (s. f.). Recuperado 28 de diciembre de 2022, de <https://www.paho.org/es/historias/atencion-pacientes-diabeticos-personas-con-problemas-salud-mental-tiempos-covid-19>
8. Ávila, S. (2022, enero 13). *En el 2021 se registraron más de 63 mil atenciones en el INADI | INADI*. <https://inadi.gob.hn/sitio/en-el-2021-se-registraron-mas-de-63-mil-atenciones-en-el-inadi/>
9. Banda, H. (2014). *Inteligencia Artificial: Principios y Aplicaciones* (1.ª ed.)

10. Bautista-Hernández, G. A., Jimenez-Nixon, D. A., & Reyes-Duke, A. M. (2022). Coral Reef Disease and Bleaching Identification through Computational Vision Algorithm. *2022 IEEE Central America and Panama Student Conference (CONESCAPAN)*, 1-6. <https://doi.org/10.1109/CONESCAPAN56456.2022.9959584>
11. Boden, M. A. (2017). *Inteligencia Artificial*. Turner.
12. Bougherara, L., Hanssens, S., Subtil, D., Vambergue, A., & Deruelle, P. (2018). Diabetes gestacional. *EMC - Ginecología-Obstetricia*, 54(1), 1-11. [https://doi.org/10.1016/S1283-081X\(18\)88086-9](https://doi.org/10.1016/S1283-081X(18)88086-9)
13. Borzouei, S., & Soltanian, A. R. (2018). Application of an artificial neural network model for diagnosing type 2 diabetes mellitus and determining the relative importance of risk factors. *Epidemiology and Health*, 40, e2018007. <https://doi.org/10.4178/epih.e2018007>
14. Bressert, E. (2012). *SciPy and NumPy: An Overview for Developers*. O'Reilly Media, Inc.
15. Campos, N., & Palomino, G. (2018). FINDRISC, utility in the screening of diabetes, personalization and associations: FINDRISC, utilidad en el screening de diabetes, personalización y asociaciones. *Revista de La Facultad de Medicina Humana*, 18(3), Art. 3. <https://doi.org/10.25176/RFMH.v18.n3.1594>
16. Carolino, I. D. R., Molena-Fernandes, C. A., Tasca, R. S., Marcon, S. S., & Cuman, R. K. N. (2008). Risk factors in patients with type 2 diabetes mellitus. *Revista Latino-Americana de Enfermagem*, 16, 238-244. <https://doi.org/10.1590/S0104-11692008000200011>
17. Carvalho, A. R. S., Guimarães, A., Werberich, G. M., de Castro, S. N., Pinto, J. S. F., Schmitt, W. R., França, M., Bozza, F. A., Guimarães, B. L. da S., Zin, W. A., & Rodrigues, R. S. (2020). COVID-19 Chest Computed Tomography to Stratify Severity and Disease Extension by Artificial Neural Network Computer-Aided Diagnosis. *Frontiers in Medicine*, 7. <https://doi.org/10.3389/fmed.2020.577609>
18. Castrillón, O. D., Sarache, W., & Castaño, E. (2017). Sistema Bayesiano para la Predicción de la Diabetes. *Información tecnológica*, 28(6), 161-168. <https://doi.org/10.4067/S0718-07642017000600017>

19. Chaki, J., Ganesh, S. T., Cidham, S. K., & S. Ananda, T. (2020, julio 4). *Machine learning and artificial intelligencebased Diabetes Mellitus detection and self-management: A systematic review* | Elsevier Enhanced Reader. <https://doi.org/10.1016/j.jksuci.2020.06.013>
20. Charbuty, B., & Abdulazeez, A. (2021). Classification Based on Decision Tree Algorithm for Machine Learning. *Journal of Applied Science and Technology Trends*, 2(01), 20-28. <https://doi.org/10.38094/jastt20165>
21. Choi, R. Y., Coyner, A. S., Kalpathy-Cramer, J., Chiang, M. F., & Campbell, J. P. (2020). Introduction to Machine Learning, Neural Networks, and Deep Learning. *Translational Vision Science & Technology*, 9(2), 14. <https://doi.org/10.1167/tvst.9.2.14>
22. *Constitución de la República de Honduras*. (1982). <https://pdba.georgetown.edu/Parties/Honduras/Leyes/constitucion.pdf>
23. Crespo, C., Brosa, M., Soria-Juan, A., Lopez-Alba, A., López-Martínez, N., & Soria, B. (2013). Costes directos de la diabetes mellitus y de sus complicaciones en España (Estudio SECCAID: Spain estimated cost Ciberdem-Cabimer in Diabetes). *Avances en Diabetología*, 29(6), 182-189. <https://doi.org/10.1016/j.avdiab.2013.07.007>
24. Cunningham, P., & Delany, S. J. (2022). k-Nearest Neighbour Classifiers—A Tutorial. *ACM Computing Surveys*, 54(6), 1-25. <https://doi.org/10.1145/3459665>
25. Curioso, W. H., Brunette, M. J., Curioso, W. H., & Brunette, M. J. (2020). Inteligencia artificial e innovación para optimizar el proceso de diagnóstico de la tuberculosis. *Revista Peruana de Medicina Experimental y Salud Publica*, 37(3), 554-558. <https://doi.org/10.17843/rpmesp.2020.373.5585>
26. Dagogo-Jack, S. (Ed.). (2017). *Diabetes Mellitus in Developing Countries and Underserved Communities*. Springer International Publishing. <https://doi.org/10.1007/978-3-319-41559-8>
27. Delgado, K., Ledesma, S., Rostro, H., Delgado, K., Ledesma, S., & Rostro, H. (2019). Análisis de electroencefalograma usando redes neuronales artificiales. *Acta universitaria*, 29. <https://doi.org/10.15174/au.2019.1672>

28. *Diabetes—OPS/OMS | Organización Panamericana de la Salud.* (s. f.). Recuperado 24 de enero de 2023, de <https://www.paho.org/es/temas/diabetes>
29. Ding, B., Qian, H., & Zhou, J. (2018). Activation functions and their characteristics in deep neural networks. 2018 *Chinese Control and Decision Conference (CCDC)*, 1836-1841. <https://doi.org/10.1109/CCDC.2018.8407425>
30. ElSayed, N. A., Aleppo, G., Aroda, V. R., Bannuru, R. R., Brown, F. M., Bruemmer, D., Collins, B. S., Hilliard, M. E., Isaacs, D., Johnson, E. L., Kahan, S., Khunti, K., Leon, J., Lyons, S. K., Perry, M. L., Prahalad, P., Pratley, R. E., Seley, J. J., Stanton, R. C., ... on behalf of the American Diabetes Association. (2022). 2. Classification and Diagnosis of Diabetes: Standards of Care in Diabetes—2023. *Diabetes Care*, 46(Supplement_1), S19-S40. <https://doi.org/10.2337/dc23-S002>
31. Enderica, P. F. V., Mendoza, Y. O. G., Apolo, K. E. M., & Flores, J. J. O. (2019). Diabetes Mellitus Tipo 2: Incidencias, Complicaciones y Tratamientos Actuales. *RECIMUNDO*, 3(1), Art. 1. [https://doi.org/10.26820/recimundo/3.\(1\).enero.2019.26-37](https://doi.org/10.26820/recimundo/3.(1).enero.2019.26-37)
32. Federación Internacional de Diabetes. (2021). *Global diabetes data report 2000—2045.* Atlas de Diabetes de La FDI. <https://diabetesatlas.org/data/>
33. Ferrer, S. M. (2020). LA ACTUACIÓN DE LA LEY DE PROTECCIÓN DE DATOS ESPAÑOLA Y EL REGLAMENTO EUROPEO: UN ESTUDIO SOBRE SU APLICACIÓN EN EL DERECHO A LA INFORMACIÓN Y EL CONSENTIMIENTO INFORMADO, DATOS CLÍNICOS Y DOCUMENTOS HOSPITALARIOS DE LOS MÉDICOS Y PACIENTES EN ESPAÑA. *Revista de Direito Brasileira*, 23(9), Art. 9. <https://doi.org/10.26668/IndexLawJournals/2358-1352/2019.v23i9.5694>
34. Gellman, M. D. (Ed.). (2020). Glycemia. En *Encyclopedia of Behavioral Medicine* (pp. 963-963). Springer International Publishing. https://doi.org/10.1007/978-3-030-39903-0_300764
35. Herman, W. H. (2017). The Global Burden of Diabetes: An Overview. En S. Dagogo-Jack (Ed.), *Diabetes Mellitus in Developing Countries and Underserved Communities* (pp. 1-5). Springer International Publishing. https://doi.org/10.1007/978-3-319-41559-8_1

36. Jaime Yelsin Rosales Malpartida. (2022). *Predicción de diabetes mellitus tipo 2 utilizando atributos médicos del Policlínico Leo SAC de San Juan de Lurigancho mediante el enfoque de Machine Learning*. 2(4), 19.
37. Jimenez Nixon, D. A. (2021). Computer vision neural network using YOLOv4 for underwater fish video detection In Roatan, Honduras. *2021 IEEE International Conference on Machine Learning and Applied Network Technologies (ICMLANT)*, 1-5.
38. Jonkman, L. J., Ndungu, M., Connor, S. E., Sharma, V. B., Pattabiraman, R., Zaver, A., Castillo, T., & Meyer, M. (2020). A qualitative assessment of diabetes care access and disease management in Central America. *Research in Social and Administrative Pharmacy*, 16(11), 1542-1549. <https://doi.org/10.1016/j.sapharm.2020.07.020>
39. Kaul, K., Tarr, J. M., Ahmad, S. I., Kohner, E. M., & Chibber, R. (2013). Introduction to Diabetes Mellitus. En S. I. Ahmad (Ed.), *Diabetes: An Old Disease, a New Insight* (pp. 1-11). Springer. https://doi.org/10.1007/978-1-4614-5441-0_1
40. Keras: Biblioteca de código abierto para crear redes neuronales. (2020, octubre 8). IONOS Digital Guide. <https://www.ionos.es/digitalguide/online-marketing/marketing-para-motores-de-busqueda/que-es-keras/>
41. Kumar, C. G., Sng, C. B. L., & Kumar, C. S. (2004). Correlation of Capillary and Venous Blood Glucometry with Laboratory Determination. *Prehospital Emergency Care*, 8(4), 378-383. <https://doi.org/10.1080/312704000905>
42. La diabetes a través de las estadísticas. (2018, agosto 17). <https://www.freestylelibre.es/libre/diabetes-blog/la-diabetes-a-traves-de-las-estadisticas.html>
43. Loncaric, F., Camara, O., Piella, G., & Bijnens, B. (2021). La integración de la inteligencia artificial en el abordaje clínico del paciente: Enfoque en la imagen cardiaca. *Revista Española de Cardiología*, 74(1), 72-80. <https://doi.org/10.1016/j.recesp.2020.07.012>
44. Lugo-Reyes, S. O., Maldonado-Colín, G., & Murata, C. (2014). Inteligencia artificial para asistir el diagnóstico clínico en medicina. *Revista Alergia México*, 61(2), 110-120. <https://doi.org/10.29262/ram.v61i2.33>

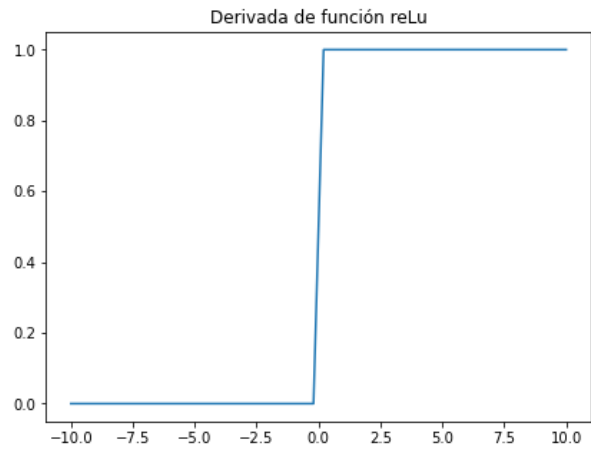
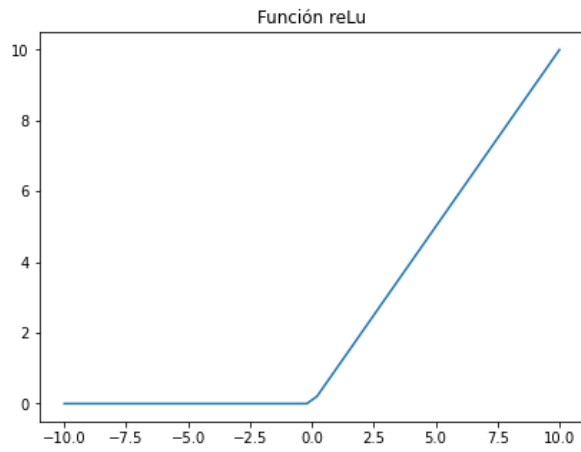
45. Martínez Bernabé, E., Paluzie-Ávila, G., Terre Ohme, S., Ruiz Poza, D., Parada Aradilla, M. A., González Martínez, J., Albertí Valmaña, R., & Castellvi Gordo, M. (2014). Sistemas de soporte a la toma de decisiones clínicas en insuficiencia renal. *Farmacia Hospitalaria*, 38(3), 216-222. <https://doi.org/10.7399/FH.2014.38.3.753>
46. *Matplotlib—Visualización con Python*. (s. f.). Recuperado 19 de febrero de 2023, de <https://matplotlib.org/>
47. Mendenhall, W., Beaver, R. J., & Beaver, B. M. (2009). *Introduction to probability and statistics* (13th ed). Brooks/Cole, Cengage Learning.
48. Microsoft 365. (s. f.). *Microsoft Forms | Surveys, Polls, and Quizzes*. Recuperado 12 de febrero de 2023, de <https://www.microsoft.com/en-us/microsoft-365/online-surveys-polls-quizzes>
49. National Cancer Institute. (2011, febrero 2). *NCI Dictionary of Cancer Terms—NCI* (nciglobal,ncienterprise) [NciAppModulePage]. <https://www.cancer.gov/publications/dictionaries/cancer-terms/def/screening>
50. OEA. (2022). *Principios Actualizados sobre la Privacidad y la Protección de Datos Personales* (1.ª ed.).
51. Ocaña-Fernández, Y., Valenzuela-Fernández, L. A., & Garro-Aburto, L. L. (2019). Inteligencia artificial y sus implicaciones en la educación superior. *Propósitos y Representaciones*, 7(2). <https://doi.org/10.20511/pyr2019.v7n2.274>
52. Olivares, R. S. (2018). Los derechos digitales en el ámbito laboral: Comentario de urgencia a la Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y Garantía de los Derechos Digitales. *IUSLabor. Revista d'anàlisi de Dret del Treball*, 3, Art. 3. <https://doi.org/10.31009/IUSLabor.2018.i03.06>
53. OPS/OMS-Secretaría de Salud de la República de Honduras. Guía de Práctica Clínica para el Manejo Ambulatorio (Promoción, Prevención, Diagnóstico y Tratamiento) del Adulto con Diabetes Mellitus Tipo 2. 2015 Diciembre. Tegucigalpa, Honduras.
54. Pablo Rodríguez de León, María Dolores Torres Soto, & Aurora Torres Soto. (2020). Mecanismo de clasificación para diabetes mellitus en la población de Aguascalientes, México. *149*, 8, 14.

55. Pariente Rodrigo, E., Deib-Morgan, K., García de Diego, O., García-Velasco, P., Sgaramella, G. A., & García González, I. (2017). Exactitud y concordancia entre glucómetros: Un estudio en condiciones habituales de práctica clínica. *Medicina de Familia. SEMERGEN*, 43(1), 20-27. <https://doi.org/10.1016/j.semerg.2016.01.011>
56. Patiño, D., Iñiguez, F., & Nivelá, M. (2022). Modelos de Machine Learning basados en Aprendizaje Supervisado para la Detección de Diabetes Mellitus en la Ciudad de Guayaquil. "Exponential Technologies and Global Challenges: Moving toward a new culture of entrepreneurship and innovation for sustainable development." <http://dx.doi.org/10.18687/LEIRD2022.1.1.208>
57. Perdomo, M. E., Cardona, M., Castro, D. M., & Mejía, W. M. (2022). Literature Review on Artificial Intelligence Implementation in the Honduran Agricultural Sector. *2022 IEEE International Conference on Machine Learning and Applied Network Technologies (ICMLANT)*, 1-5.
58. Pineda, J. M. (2022). Modelos predictivos en salud basados en aprendizaje de maquina (machine learning). *Revista Médica Clínica Las Condes*, 33(6), 583-590. <https://doi.org/10.1016/j.rmclc.2022.11.002>
59. Poveda, K. A. F., García, K. J. Q., Subía, D. L. F., & Choez, C. A. C. (2020). Utilidad de hemoglobina glicosilada en diabetes tipo 2. *RECIAMUC*, 4(3), Art. 3. [https://doi.org/10.26820/reciamuc/4.\(3\).julio.2020.118-126](https://doi.org/10.26820/reciamuc/4.(3).julio.2020.118-126)
60. Richesson, R. L., Rusincovitch, S. A., Wixted, D., Batch, B. C., Feinglos, M. N., Miranda, M. L., Hammond, W. E., Califf, R. M., & Spratt, S. E. (2013). A comparison of phenotype definitions for diabetes mellitus. *Journal of the American Medical Informatics Association*, 20(e2), e319-e326. <https://doi.org/10.1136/amiajnl-2013-001952>
61. Ríos, B. R. (2022). Diagnóstico del dengue utilizando redes neuronales artificiales. *Ciencia Latina Revista Científica Multidisciplinar*, 6(4), Art. 4. https://doi.org/10.37811/cl_rcm.v6i4.3037
62. Röder, P. V., Wu, B., Liu, Y., & Han, W. (2016). Pancreatic regulation of glucose homeostasis. *Experimental & Molecular Medicine*, 48(3), e219. <https://doi.org/10.1038/emm.2016.6>

63. Sampieri, R. H. (2014). *Metodología de la Investigación* (6ta Edición). McGraw-Hill Education.
64. Sánchez, L., & Núñez, G. (2019). Diagnóstico de Enfermedades Endémicas con redes neuronales artificiales a través WEKA. *Ciencia Latina Revista Científica Multidisciplinar*, 3(1), Article 1.
65. Sarmiento-Ramos, J. L. (2020). Aplicaciones de las redes neuronales y el deep learning a la ingeniería biomédica. *Revista UIS Ingenierías*, 19(4), 1-18. <https://doi.org/10.18273/revuin.v19n4-2020001>
66. Schnell, O., Barnard, K., Bergenstal, R., Bosi, E., Garg, S., Guerci, B., Haak, T., Hirsch, I. B., Ji, L., Joshi, S. R., Kamp, M., Laffel, L., Mathieu, C., Polonsky, W. H., Snoek, F., & Home, P. (2015). Clinical Utility of SMBG: Recommendations on the Use and Reporting of SMBG in Clinical Research. *Diabetes Care*, 38(9), 1627-1633. <https://doi.org/10.2337/dc14-2919>
67. Serna, E. (2017). Principios y características de las redes neuronales artificiales. En *DESARROLLO E INNOVACIÓN EN INGENIERÍA* (2.ª ed., pp. 212-225). Editorial Instituto Antioqueño de Investigación.
68. Scikit-learn: Machine learning in Python—Scikit-learn 1.2.2 documentation. (s. f.). Recuperado 24 de marzo de 2023, de <https://scikit-learn.org/stable/>
69. Sprockel, J., Diaztagle, J., Alzate, W., & González, E. (2014). Redes neuronales en el diagnóstico del infarto agudo de miocardio. *Revista Colombiana de Cardiología*, 21. <https://doi.org/10.1016/j.rccar.2013.10.001>
70. Software de hojas de cálculo Microsoft Excel | Microsoft 365. (s. f.). Recuperado 22 de febrero de 2023, de <https://www.microsoft.com/es/microsoft-365/excel>
71. TensorFlow. (2020). <https://www.tensorflow.org/?hl=es-419>
72. tkinter—Python interface to Tcl/Tk. (2023). Python documentation. <https://docs.python.org/3/library/tkinter.html>
73. Vega, M. Á., Mora, L. M. Q., & Badilla, M. V. C. (2020). Inteligencia artificial y aprendizaje automático en medicina. *Revista Médica Sinergia*, 5(8), Art. 8. <https://doi.org/10.31434/rms.v5i8.557>

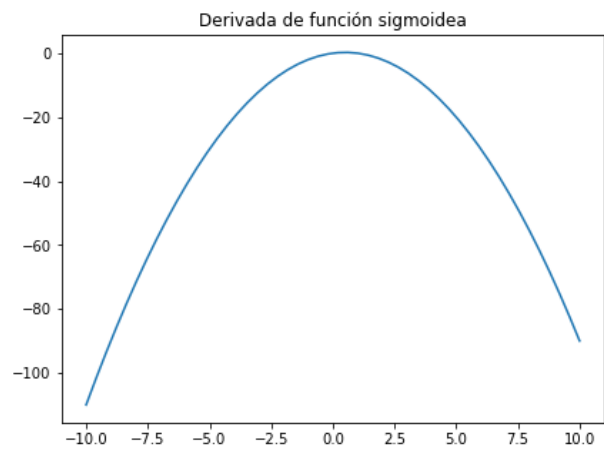
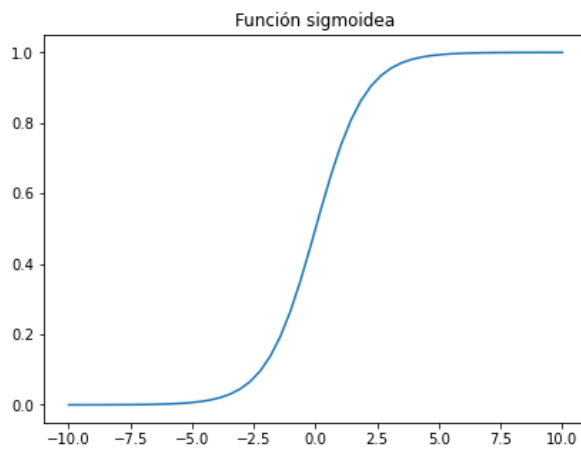
74. Visión artificial, OpenCV y Python. (2016, mayo 23). primeros pasos para analizar imágenes. <https://programarfacil.com/podcast/81-vision-artificial-opencv-phyton/>
75. Visual Studio: IDE and Code Editor for Software Developers and Teams. (s. f.). Visual Studio. Recuperado 3 de marzo de 2023, de <https://visualstudio.microsoft.com>
76. Wackerly, Dennis D, William Mendenhall III, & Richard L. Scheaffer. (s. f.). *Estadística matemática con aplicaciones* (séptima). Cengage Learning Editores, <http://latinoamerica.cengage.com>
77. Walpole, R. E., Myers, R. H., Myers, S. L., & Ye, K. (2012). *Probabilidad y estadística para ingeniería y ciencias* (novena). Pearson Educación. https://bibliotecavirtualaserena.files.wordpress.com/2017/05/libro_probabilidad-y-estadistica-para-ingenerc3ada-y-ciencias-ronald-e-walpole-mayers.pdf
78. Zajmi, L., Ahmed, F. Y. H., & Jaharadak, A. A. (2018). Concepts, Methods, and Performances of Particle Swarm Optimization, Backpropagation, and Neural Networks. *Applied Computational Intelligence and Soft Computing*, 2018, e9547212. <https://doi.org/10.1155/2018/9547212>
79. Zhang, Z. (2016). Introduction to machine learning: K-nearest neighbors. *Annals of Translational Medicine*, 4(11), 218. <https://doi.org/10.21037/atm.2016.03.37>

ANEXOS



Anexo 1 Función de activación utilizada en la capa de entrada.

Fuente: Elaboración propia.



Anexo 2 Función de activación utilizada en la capa oculta.

Fuente: Elaboración propia.

Anexo 3 Base de datos de prueba.

No.	Edad	Peso	Sexo	Glucosa en Sangre	Hemoglobina glicosilada	Colesterol	Presión arterial sistólica	Presión arterial diastólica	Triglicéridos	Creatinina
1	22	128	1	126	6.5	136.7	110	70	51	0.8
2	27	145	1	101	5	229.62	160	90	56.71	0.7
3	70	173	1	94	11.3	193.2	139	86	218.6	0.67
4	71	212	2	123	6.1	152	110	80	142	1.56
5	58	130	1	107	6.5	167	237	105	162	0.65
6	61	162	1	179	7.8	269	160	80	266	1.2
7	45	255	1	168	7.2	202.6	130	70	234	0.7
8	61	146	1	108	6.2	121	150	90	162.9	1.03
9	59	134	1	300	7.5	228	130	70	840	1.3
10	52	144	1	263	7.59	231.4	142	101	103.5	0.9
11	55	144.98	1	100	8.5	264	130	70	238	0.8
12	60	150	1	199	8	220	128	67	137	0.69
13	65	126	1	362	10	205	158	67	430	1.1
14	78	160	2	150	7	185	180	60	198	0.71
15	77	146	2	103	5.9	224	160	90	174	1.07
16	70	154	1	98.3	8.25	270.1	140	80	489.9	1.3
17	34	158	2	70.4	5.2	158	124	88	106	0.98
18	66	133	2	93	4.9	158	140	90	52	1
19	77	158	1	127	6.65	286	223	109	203	1.02
20	48	243	2	263	6.57	151	110	90	360	0.79
21	56	147	2	357	7.2	262.2	120	90	113.1	0.5
22	56	128	1	210	9.1	209	132	72	113.4	1.3
23	53	143	1	396.8	13.54	289.7	120	100	370.8	1.3
24	38	126	1	66.2	15.4	194	120	76	144	1.48
25	89	132.5	2	231	6.7	165	152	90	73	1.2
26	83	143	2	143	6.9	231	130	70	137	0.5
27	42	152	2	138	6.4	128	100	60	205.8	1.1

No.	Edad	Peso	Sexo	Glucosa en Sangre	Hemoglobina glicosilada	Colesterol	Presión arterial sistólica	Presión arterial diastólica	Triglicéridos	Creatinina
28	74	162	2	98	5.4	205	120	80	381	1.1
29	43	164	1	156	6.9	140	137	80	140	0.84
30	67	169	1	164	7.21	164	148	90	131	1
31	33	190	2	171	6.8	118.9	110	70	214.8	0.84
32	88	116	1	94	5.5	118.9	125	182	145	0.78
33	55	246	1	111.1	5.3	168	122	73	71	0.83
34	61	174	1	113	5.4	184	110	70	148	0.9

Fuente: Expedientes médicos de pacientes de Hospital El Progreso.