

UNIVERSIDAD NACIONAL DE JAÉN
FACULTAD DE CIENCIAS DE LA SALUD
CARRERA PROFESIONAL DE TECNOLOGÍA MÉDICA



**UNIVERSIDAD NACIONAL
DE JAÉN**

**PROPUESTA DE UN MODELO DE RED NEURONAL
ARTIFICIAL PARA PREDECIR DIABETES MELLITUS EN LA
CIUDAD DE JAÉN-2023**

**INFORME FINAL DE TESIS PARA OPTAR EL TÍTULO
PROFESIONAL DE LICENCIADO TECNÓLOGO MÉDICO EN
LABORATORIO CLÍNICO Y ANATOMÍA PATOLÓGICA**

AUTORAS

Bach. Clara Cerna Reátegui

Bach. Evelyn Anyelí Molocho Martínez

ASESORES

Dr. Nilthon Arce Fernández

Dr. Guillermo Núñez Sánchez

LÍNEA DE INVESTIGACIÓN

Enfermedades No Transmisibles

JAÉN – PERÚ

2026

Clara Cerna Reátegui Evelyn Anyelí Molocho Martín...

PROPUESTA DE UN MODELO DE RED NEURONAL ARTIFICIAL PARA PREDECIR DIABETES MELLITUS EN LA CIUDAD DE JAÉ...

- PROYECTOS DE TESIS E INFORMES FINALES 2026
- Proyectos e Informes en evaluación
- Universidad Nacional de Jaén

Detalles del documento

Identificador de la entrega

trn:oid:::1:3516115662

88 páginas

Fecha de entrega

24 mar 2026, 11:40 a.m. GMT-5

22.446 palabras

Fecha de descarga

24 mar 2026, 11:48 a.m. GMT-5

128.658 caracteres

Nombre del archivo

INFORME_FINAL_DE_TESIS-CLARA_CERNA_REATEGUI_Y_EVELYN_MOLOCHO_MARTINEZ_-TM-O-20....docx

Tamaño del archivo

1.5 MB

 UNIVERSIDAD NACIONAL DE JAÉN

Dr. Guillermo Nuñez Sánchez
RESPONSABLE DE LA UNIDAD DE INVESTIGACIÓN
DE LA FACULTAD DE CIENCIAS DE LA SALUD




6% Similitud general

El total combinado de todas las coincidencias, incluidas las fuentes superpuestas, para ca...

Filtrado desde el informe


- Bibliografía
- Texto citado
- Texto mencionado
- Coincidencias menores (menos de 12 palabras)

Fuentes principales

- 4%  Fuentes de Internet
- 1%  Publicaciones
- 5%  Trabajos entregados (trabajos del estudiante)

Marcas de integridad


N.º de alerta de integridad para revisión

-  **Caracteres reemplazados**
72 caracteres sospechosos en N.º de páginas
Las letras son intercambiadas por caracteres similares de otro alfabeto.

Los algoritmos de nuestro sistema analizan un documento en profundidad para buscar inconsistencias que permitirían distinguirlo de una entrega normal. Si advertimos algo extraño, lo marcamos como una alerta para que pueda revisarlo.

Una marca de alerta no es necesariamente un indicador de problemas. Sin embargo, recomendamos que preste atención y la revise.

 UNIVERSIDAD NACIONAL DE JAÉN


Dr. Guillermo Nuñez Sánchez
RESPONSABLE DE LA UNIDAD DE INVESTIGACIÓN
DE LA FACULTAD DE CIENCIAS DE LA SALUD



UNIVERSIDAD NACIONAL DE JAÉN

Ley de Creación N° 29304

Universidad Licenciada con Resolución del Consejo Directivo N° 002-2018-SUNEDU /CD

ACTA DE SUSTENTACIÓN

En la ciudad de Jaén, el martes 10 de marzo de 2026, siendo las 16:00 horas, se reunieron los integrantes del Jurado:

Presidente : **Dr. Diomer Marino Jara Llanos.**
Secretario : **Dr. Juan Enrique Arellano Ubillus.**
Vocal : **Mg. Frans Fuentes Maza.**

Para evaluar la Sustentación del Informe Final de:

- () Trabajo de Investigación
() Tesis
() Trabajo de Suficiencia Profesional

Titulado: **"PROPUESTA DE UN MODELO DE RED NEURONAL ARTIFICIAL PARA PREDECIR DIABETES MELLITUS EN LA CIUDAD DE JAÉN-2023"** presentado por las bachilleres **Clara Cerna Reátegui y Evelyn Anyelí Molocho Martínez**, de la Escuela Profesional de Tecnología Médica de la Universidad Nacional de Jaén.

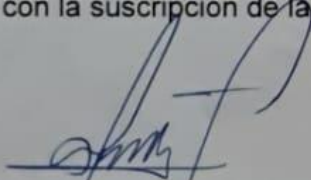
Después de la sustentación y defensa, el Jurado acuerda:

- () Aprobar () Desaprobar () Unanimidad () Mayoría

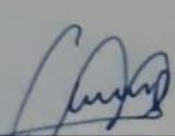
Con la siguiente mención:

- | | | |
|---------------------------|------------|---|
| a) Excelente | 18, 19, 20 | () |
| b) Muy bueno | 16, 17 | () |
| c) Bueno | 14, 15 | () |
| d) Regular | 13 | (<input checked="" type="checkbox"/>) |
| e) Desaprobado 12 o menos | | () |

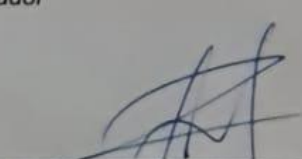
Siendo las 17:00 horas del mismo día, el Jurado concluye el acto de sustentación confirmando su participación con la suscripción de la presente.



Dr. Diomer Marino Jara Llanos.
Presidente Jurado Evaluador



Dr. Juan Enrique Arellano Ubillus.
Secretario Jurado Evaluador



Mg. Frans Fuentes Maza.
Vocal Jurado Evaluador

“Año de la Esperanza y el Fortalecimiento de la Democracia”

ANEXO N°06:

DECLARACIÓN JURADA DE AUTENTICIDAD Y DE NO PLAGIO DE LA TESIS O TRABAJO DE INVESTIGACIÓN (PREGRADO)

Yo, **Clara Cerna Reátegui y Evelyn Anyelí Molocho Martínez**, egresado de la carrera Profesional de **Tecnología Médica** de la Facultad de Ciencias de la salud de la Universidad Nacional de Jaén, identificado (a) con DNI **75468866** y **71869858** respectivamente.

Declaro bajo juramento que:

1. Soy Autor del trabajo titulado:

“PROPUESTA DE UN MODELO DE RED NEURONAL ARTIFICIAL PARA PREDECIR DIABETES MELLITUS EN LA CIUDAD DE JAÉN-2023

Asesorado por: **Dr. Nilthon Arce Fernández y Dr. Guillermo Núñez Sánchez.**

El mismo que presento bajo la modalidad de sustentación para optar; el Título Profesional/Grado Académico de Título profesional de: Licenciado Tecnólogo Médico en Laboratorio Clínico y Anatomía Patológica.

2. El texto de mi trabajo final respeta y no vulnera los derechos de terceros, incluidos los derechos de propiedad intelectual. En el sentido, el texto de mi trabajo final no ha sido plagiado total ni parcialmente, para la cual he respetado las normas internacionales de citas y referencias de las fuentes consultadas.
3. El texto del trabajo final que presento no ha sido publicado ni presentado antes en cualquier medio electrónico o físico.
4. La investigación, los resultados, datos, conclusiones y demás información presentada que atribuyo a mi autoría son veraces.
5. Declaro que mi trabajo final cumple con todas las normas de la Universidad Nacional de Jaén.
6. Soy consciente de que el hecho de no respetar los derechos de autor y hacer plagio, es objeto de sanciones universitarias y/o legales.

El incumplimiento de lo declarado da lugar a responsabilidad del declarante, en consecuencia; a través del presente documento asumo frente a terceros, la Universidad Nacional de Jaén y/o la Administración Pública toda responsabilidad que pueda derivarse por el trabajo final presentado. Lo señalado incluye responsabilidad pecuniaria incluido el pago de multas u otros por los daños y perjuicios que se ocasionen.

Fecha: Jaén, 10 de junio del 2026.



Clara Cerna Reátegui.



Evelyn Anyelí Molocho Martínez

INDICE

ÍNDICE DE TABLAS	iv
ÍNDICE DE FIGURAS	v
RESUMEN	vi
ABSTRACT	vii
I. INTRODUCCIÓN	7
II. MATERIAL Y MÉTODOS	17
III. RESULTADOS	26
IV. DISCUSIÓN	32
V. CONCLUSIONES Y RECOMENDACIONES	35
VI. REFERENCIAS BIBLIOGRAFICAS	37
AGRADECIMIENTO	43
DEDICATORIA	44
ANEXOS	45

ÍNDICE DE TABLAS

Tabla 1. Datos de diagnóstico clínicos relevantes sobre diabetes mellitus	26
Tabla 2. Registro de datos observados del Hospital – San Javier de Bellavista, ciudad de Jaén. ...	55
Tabla 3. Estadísticos descriptivos de los datos en Edad.....	58
Tabla 4. Estadísticos descriptivos de los datos del IMC.	59
Tabla 5. Estadísticos descriptivos de los datos del PAM.	60
Tabla 6. Estadísticos descriptivos de los datos de la Glicemia Basal.....	61
Tabla 7. Estadísticos descriptivos de los datos de la Hemoglobina Glicosilada.	62
Tabla 8. Estadísticos descriptivos de los datos de la Tolerancia Oral a la Glucosa.	63
Tabla 9. Estudio de correlación de variables mediante Pearson.....	65
Tabla 10. Registro de datos de entrada del Hospital – San Javier de Bellavista, ciudad de Jaén....	79
Tabla 11. Resultados de la predicción de diabetes mellitus.	89

ÍNDICE DE FIGURAS

Figura 1. Mapa de calor para determinar la relevancia de las variables.	28
Figura 2. Validación gráfica del modelo de red neuronal artificial.....	30
Figura 3. Secuencia del proceso CRISP-DM. Adaptado de Galán Cotrina V (UC3M). ⁽⁴⁶⁾	50
Figura 4. Estadígrafo del registro de datos observados del Hospital San Javier de Bellavista, ciudad de Jaén.	57
Figura 5. Evolución de la función de pérdida durante el entrenamiento y validación de la red neuronal.....	71
Figura 6. Diagrama de Flujo del proceso de predicción del estado de Diabetes a partir de un nuevo paciente mediante un modelo predictivo.....	74
Figura 7. Funcionamiento lógico de la red neuronal artificial	76
Figura 8. Código en Python para calcular las métricas de evaluación	78
Figura 9. Estadígrafo del registro de datos pronósticos del Hospital San Javier de Bellavista, ciudad de Jaén.	81
Figura 10. Diagrama de dispersión de valores de IMC.....	84
Figura 11. Diagrama de dispersión de valores del PAM.....	85
Figura 12. Diagrama de dispersión de valores de la Glicemia.	86
Figura 13. Diagrama de dispersión de valores HbA1c [%].....	87
Figura 14. Diagrama de dispersión de valores de la Tolerancia Oral a la Glucosa [mg/dL].	88

RESUMEN

La investigación “Propuesta de un modelo de red neuronal artificial para predecir diabetes mellitus en la ciudad de Jaén–2023” tuvo como objetivo desarrollar un método de diagnóstico con capacidades clínicas probabilísticas mediante red LSTM implementada en Python. El estudio se realizó con datos clínicos y de laboratorio del Hospital San Javier de Bellavista. Se trabajó con 415 historias clínicas de pacientes entre 18 a 70 años, seleccionadas por muestreo no probabilístico por conveniencia. Enfoque cuantitativo, aplicado y explicativo, diseño no experimental y transversal, método deductivo y la metodología CRISP-DM guiaron el proceso analítico. Considerándose; edad, género, IMC, presión arterial, glucosa basal, HbA1c y prueba de tolerancia oral a la glucosa (OGTT). El modelo LSTM mostró reducción sostenida de la función de pérdida a lo largo de 600 épocas, con estabilidad desde la época 150 y sin sobreajuste (curvas de entrenamiento y validación próximas). Los principales predictores fueron HbA1c ($r=0.718$), OGTT ($r=0.643$) y glucosa basal ($r=0.586$); los promedios observados fueron 131.75 mg/dL (glucosa) y 6.47% (HbA1c), sugiriendo alta proporción de prediabetes/DM2. El rendimiento fue comparable a reportes internacionales (77–91%), evidenciando potencial de aplicabilidad clínica para el tamizaje y diagnóstico temprano en contextos con recursos limitados como Jaén.

Palabras clave: diabetes mellitus; red neuronal LSTM; predicción clínica.

ABSTRACT

The research project “Proposal of an artificial neural network model to predict diabetes mellitus in the city of Jaén–2023” aimed to develop a diagnostic method with probabilistic clinical capabilities using an LSTM network implemented in Python. The study was conducted using clinical and laboratory data from the San Javier de Bellavista Hospital. It analyzed 415 medical records of patients aged 18 to 70 years, selected by non-probability convenience sampling. A quantitative, applied, and explanatory approach, a non-experimental and cross-sectional design, a deductive method, and the CRISP-DM methodology guided the analytical process. The following variables were considered: age, gender, BMI, blood pressure, fasting glucose, HbA1c, and oral glucose tolerance test (OGTT). The LSTM model showed a sustained reduction in the loss function over 600 epochs, with stability from epoch 150 onward and no overfitting (training and validation curves close together). The main predictors were HbA1c ($r = 0.718$), OGTT ($r = 0.643$), and fasting glucose ($r = 0.586$); the observed averages were 131.75 mg/dL (glucose) and 6.47% (HbA1c), suggesting a high proportion of prediabetes/T2DM. The performance was comparable to international reports (77–91%), demonstrating potential clinical applicability for screening and early diagnosis in resource-limited settings such as Jaén.

Keywords: diabetes mellitus; LSTM neural network; clinical prediction.

I. INTRODUCCIÓN

La diabetes mellitus, es una de las amenazas más grandes para la salud pública a nivel global, por su constante aumento afecta tanto a naciones avanzadas como en desarrollo. Esto impacta en el cumplimiento del Objetivo de Desarrollo Sostenible 3 (Salud y Bienestar), busca asegurar una vida saludable y fomentar el bienestar en todas las etapas de la vida. Se estimaba que alrededor de 537 millones de adultos padecían diabetes, en el año 2021. Se proyecta que para el 2030 alcance 643 millones y a los 783 millones en el año 2045 ^(1,2).

Esta enfermedad causó 6,7 millones de fallecidos, en el 2021. Convirtiéndose a nivel mundial en la octava causa de muerte. Dentro de este marco, el objetivo 3 y 4 del Objetivo de Desarrollo Sostenible propone disminuir en un tercio de la mortalidad prematura vinculada a enfermedades no transmisibles, a través de tratamientos y medidas preventivas. Resulta fundamental incorporar herramientas tecnológicas predictivas, para alcanzar esta meta. La implementación de modelos de inteligencia artificial constituye una estrategia innovadora para anticipar casos, facilitar el diagnóstico y disminuir la carga socioeconómica. En la provincia de Jaén, donde los recursos médicos son escasos, esta tecnología podría marcar un cambio significativo en la prevención de complicaciones severas ⁽¹⁾.

La Diabetes Mellitus ha experimentado un aumento notable a escala global, especialmente en países de ingresos bajos y medios. Aproximadamente el 81% de los adultos con diabetes (equivalente a 4 de cada 5 personas), viven en estos países. En 2021, esto representó el 75 % de los casos registrados, lo cual evidencia una clara desigualdad en el acceso a diagnósticos tempranos y tratamientos efectivos, esto según el International Diabetes Federation (IDF) ⁽¹⁾.

Destaca en estas cifras Asia, China e India, acumulando más del 40% de los casos a nivel mundial en 2021. Un desafío crítico es la carencia de detección temprana, se estima que 1 de cada 2 adultos con diabetes no ha sido diagnosticado. Esto conlleva a graves consecuencias: aumentan las complicaciones, costos de atención que se duplican y la esperanza de vida disminuye. La Inteligencia Artificial (IA) se presenta como una herramienta efectiva para reducir errores humanos, frente a esta situación crítica, acelerar los diagnósticos y optimizar la gestión sanitaria, especialmente en contextos con escasos recursos ⁽¹⁾

La investigación realizada en Pakistán en el año 2020, que consistió en un análisis literario sobre cómo la Inteligencia Artificial puede ser utilizada para detectar la diabetes, abarcando diferentes enfoques a partir de más de 35 publicaciones, establece que la Red Neuronal Profunda, al usar los mismos datos para construcción y pruebas, muestra un desempeño mejorado, aunque implica un mayor tiempo de procesamiento y ajustes regulares⁽³⁾.

En una investigación llevada a cabo en la India en 2023, se desarrolló un sistema de atención médica que utilizaba tecnología de visión por computadora para detectar diabetes a través de inteligencia artificial. Implementaron un modelo con máquinas de vectores de soporte (SVM) logrando una precisión del 94 % en el diagnóstico de diabetes. Además, utilizaron una red neuronal convolucional (CNN) que tuvo un 80% de efectividad en reconocer características de la retinopatía diabética en imágenes de la retina. A pesar de que la CNN presenta algunas limitaciones, como la identificación de ciertos aspectos de la retinopatía diabética, es útil para distinguir entre retinas sanas y las que están afectadas ⁽⁴⁾.

La investigación desarrollada en China en 2023 presenta un sistema de diagnóstico para la diabetes, empleando algoritmos de aprendizaje automático. Se utilizaron 768 registros de datos de mujeres entre 21 y 81 años, en el análisis de los datos resalta la glucosa como un factor esencial, además de la edad y el IMC. Logrando una precisión del 94.67 % en la predicción de la enfermedad ⁽⁵⁾.

En otro contexto, una investigación llevada a cabo en China en 2024 presenta un enfoque para diagnosticar diabetes mellitus utilizando ultrasonido de alta frecuencia y una red neuronal convolucional. Al examinar la correlación entre la glucosa y la hemoglobina glucosilada en las células sanguíneas, se expusieron células a distintas concentraciones de glucosa y se registraron señales acústicas utilizando un transductor de 90 MHz, analizadas a través de la red neuronal. Esta red fue capaz de detectar variaciones en las características de los glóbulos rojos causadas por la glucosa, alcanzando una tasa de clasificación del 98 %. Esta innovación sugiere un método de diagnóstico no invasivo en vivo, eludiendo la necesidad de muestras de sangre gracias a la integración de ultrasonido de alta frecuencia y redes neuronales ⁽⁶⁾.

En Antioquia, se llevó a cabo la investigación titulada “Uso de mapas cognitivos difusos junto con algoritmos genéticos para anticipar la diabetes mellitus tipo 2” durante el

año 2024, donde se evaluaron 768 individuos del grupo indígena PIMA, logrando un nivel de precisión del 99 %, una exactitud del 98% y un recall del 100% en la anticipación de la diabetes mellitus tipo 2. Este modelo, debido a su capacidad predictora podría ser útil como herramienta de apoyo en la detección temprana de la enfermedad y en la elaboración de decisiones clínicas⁽⁷⁾.

En una investigación desarrollada en Ecuador durante el año 2023 sobre los algoritmos de inteligencia artificial más comunes para el diagnóstico de la diabetes mellitus tipo 2, se observó que la Red Neuronal Artificial (ANN) destacó por su capacidad de procesamiento, mientras que el SVM mostró un buen desempeño con un menor número de datos. Para este análisis se consideraron variables como el IMC, el nivel de glucosa y la edad, sugiriendo que incluir más variables podría aumentar la exactitud del diagnóstico⁽⁸⁾.

En el año 2023, en Honduras, se desarrolló una investigación cuyo objetivo fue diagnosticar la enfermedad Diabetes Mellitus. Se recopilaron datos de 170 pacientes que fueron procesados en el lenguaje de programación Python, y luego utilizados para desarrollar un modelo de red neuronal artificial con un nivel de precisión del 91%, demostrando efectividad en la predicción de la enfermedad⁽⁹⁾.

La tasa de Diabetes Mellitus tipo 2 ha mostrado un crecimiento continuo, en Perú en el año 2023. Afectando al 7,1 % de los adultos de acuerdo con el informe de la Encuesta Nacional de Salud y Nutrición (ENSANUT), no obstante, este dato se eleva a 12,4 % en áreas urbanas como Lima Metropolitana⁽¹⁰⁾, la situación es preocupante, ya que alrededor del 41.3 % de los casos de diabetes mellitus tipo 2 en Perú no están diagnosticados, lo que incrementa el riesgo de complicaciones como ceguera, insuficiencia renal o amputaciones. Desde el punto de vista económico, el tratamiento de la diabetes y sus consecuencias representan una carga significativa para el sistema de salud peruano, llegando a consumir aproximadamente el 30 % del presupuesto nacional de salud, según un estudio reciente sobre el costo de la enfermedad^(11,12). En este contexto, el desarrollo de modelos predictivos basados en IA puede contribuir significativamente a la identificación temprana de personas en riesgo, optimizando el uso de los recursos disponibles y fortaleciendo las estrategias de prevención en el primer nivel de atención.

Un estudio ejecutado en el año 2023, en Lima – Perú, se implementaron diversas tecnologías de aprendizaje supervisado, utilizando el programa WEKA como herramienta

principal de análisis para predecir la diabetes. Presentaron un modelo híbrido que logró 73.03 % de precisión y otro similar que alcanzó 58.3 % en especificidad, 86.8 % en sensibilidad y 77 % en exactitud. Se destacó la importancia de las variables relevantes y enfatizaron la necesidad de abordar el desequilibrio en los datos para mejorar la precisión de los modelos desarrollados ⁽²⁾.

En el norte del país, destaca una investigación llevada a cabo en Chiclayo durante el año 2022. El propósito central de este estudio fue desarrollar un modelo de red neuronal convolucional (CNN) orientada al diagnosticar la diabetes. Para lograr este objetivo, procesaron 516 imágenes de fondo de ojo; de estas, destinaron 413 para el entrenamiento y los 103 restantes se utilizaron para la prueba. El modelo alcanzó en precisión 83.50 %, sensibilidad 85.50 % y especificidad 79.41 %. Además, mostró un rendimiento superior en comparación con otros modelos desarrollados ⁽¹³⁾.

Asimismo, otro estudio desarrollado en el 2023 en un Centro de Salud del distrito de Jaén reportó que el 38 % de las mujeres evaluadas tenían niveles elevados de glucosa basal, lo que refuerza la necesidad de mejorar las estrategias de prevención y diagnóstico temprano en esta zona ⁽¹⁴⁾. La falta de campañas preventivas efectivas y la escasa infraestructura para realizar pruebas en áreas periféricas obstaculizan una respuesta adecuada. Por ello, la implementación de un modelo basado en redes neuronales artificiales podría facilitar la identificación precisa de individuos con alto riesgo de desarrollar diabetes, mejorando así la cobertura y la efectividad del sistema de salud.

En la provincia de Jaén, se ha evidenciado una tendencia creciente en los casos de Diabetes Mellitus tipo 2, especialmente en adultos mayores de 35 años. Diversos estudios realizados en centros de salud de la región muestran una mayor incidencia en mujeres y una estrecha relación con el sobrepeso y la obesidad. Una investigación realizada en el 2022 en el Centro de Salud Nueva Esperanza, en Jaén, se encontró que el 66% de los casos diagnosticados correspondían a mujeres, mientras que el 34 % presentaba obesidad o sobrepeso como condición asociada ⁽¹⁵⁾.

Es importante mencionar que, en la búsqueda del estado del arte a nivel local, no se ha encontrado ninguna investigación relacionada con el desarrollo de un modelo de red neuronal artificial para predecir la enfermedad no transmisible diabetes mellitus.

El aumento constante de la diabetes mellitus en zonas urbanas y rurales de Perú, representa un grave riesgo para el bienestar de la población. En la región Cajamarca, donde se localiza la provincia de Jaén, los servicios de salud aún enfrentan desafíos en cuanto al diagnóstico temprano, lo que incrementa el peso económico y social del tratamiento en etapas avanzadas. Por ello, considerando las recomendaciones de la Organización Mundial de la Salud y la Organización Panamericana de la Salud, que sugieren la adopción de tecnologías digitales para mejorar la vigilancia epidemiológica y dirigir las intervenciones preventivas, se desarrolló un modelo predictivo fundamentado en inteligencia artificial que funcionará como una herramienta para dar prioridad a la atención de los grupos más vulnerables.

Esta investigación tiene como bases a los estudios de: Liu et al.⁽¹⁶⁾, llevó a cabo en 2024 en Corea, la investigación denominada *“Noninvasive Detection of Diabetes Mellitus Using High-Frequency Ultrasound and Convolutional Neural Networks”*, cuyo objetivo fue desarrollar un método no invasivo para el diagnóstico de la diabetes. La metodología incluyó el uso de ultrasonido de alta frecuencia y redes neuronales convolucionales para analizar cambios en los glóbulos rojos. Los resultados alcanzaron una precisión cercana al 98%. Se concluye que esta técnica representa una alternativa prometedora al diagnóstico tradicional invasivo.

En esa misma línea de investigación, el estudio de Hoyos W.⁽⁷⁾, en Colombia durante el 2024, denominado *“Uso de mapas cognitivos difusos junto con algoritmos genéticos para anticipar la diabetes mellitus tipo 2”*. Su principal objetivo estuvo ligado a diseñar un modelo predictivo para la detección temprana de la diabetes tipo 2. Destacando su enfoque al combinar mapas cognitivos difusos con algoritmo genéticos bajo un diseño cuantitativo y experimental. Al evaluar el rendimiento de esta herramienta, alcanzaron niveles sobresalientes tanto en exactitud como en sensibilidad. En definitiva, el modelo propuesto es una herramienta de apoyo a la toma de decisiones clínicas.

Por otro lado, se ejecutó un estudio realizado en Estados Unidos durante el 2023 por Zhang et al.⁽¹⁷⁾. Bajo el título: *“Application of Machine Learning Algorithms in Diabetes Diagnosis Based on Clinical Data”*, donde los autores buscaban medir el desempeño clínico de diferentes algoritmos de aprendizaje automático en el diagnóstico de la diabetes mellitus a partir de datos clínicos. Al aplicar un enfoque cuantitativo con diseño experimental, el equipo no se limitó a una sola herramienta, más bien, analizaron en conjunto cómo

respondían la regresión logística, las máquinas de soporte vectorial, los árboles de decisión y las redes neuronales frente a los mismos historiales médicos. Los datos clínicos fueron procesados y normalizados antes del entrenamiento de los modelos. Los resultados evidenciaron que los modelos basados en aprendizaje automático alcanzaron altos niveles de precisión diagnóstica, destacando las redes neuronales y los modelos ensemble. Se concluye que el aprendizaje automático representa una herramienta eficaz para apoyar el diagnóstico clínico de la diabetes, especialmente cuando se dispone de grandes volúmenes de datos clínicos.

Además, Reinoso et al. ⁽¹⁸⁾ realizaron en Ecuador el año 2023 el estudio *“Evaluación de algoritmos de inteligencia artificial para el diagnóstico de la diabetes mellitus tipo 2”*, cuyo objetivo fue comparar distintos algoritmos de IA. Esta metodología fue cuantitativa y comparativa, aplicando varios modelos de clasificación. Los resultados evidenciaron que los algoritmos de aprendizaje automático mejoran el rendimiento diagnóstico frente a métodos tradicionales. Se concluye que la inteligencia artificial tiene un alto potencial en el diagnóstico de la diabetes tipo 2.

Del mismo modo, Ahmad et al. ⁽¹⁹⁾, en Pakistán en el año 2023, realizaron el estudio *“A Deep Learning Model for Early Detection of Diabetes Using Neural Networks”*, con el objetivo de diseñar un modelo de aprendizaje profundo para la detección temprana de la diabetes. La metodología fue cuantitativa y experimental, empleando redes neuronales profundas entrenadas con datos clínicos. Los resultados evidenciaron una alta precisión diagnóstica. Los autores concluyen que los modelos de deep learning pueden mejorar significativamente la detección temprana de la diabetes.

Asimismo, en Perú, Huamán et al ⁽²⁰⁾. realizaron en 2023 el estudio *“Implementación de modelos híbridos de predicción de diabetes con aprendizaje supervisado en WEKA”*, cuyo objetivo fue evaluar modelos híbridos utilizando el software WEKA. La metodología consistió en aprendizaje supervisado con técnicas combinadas. Los resultados mostraron mejoras en la precisión al aplicar preprocesamiento de datos. Se concluye que los modelos híbridos representan una alternativa viable para la predicción de la diabetes.

De igual manera, Pineda et al. ⁽²¹⁾, en 2023, desarrollaron el estudio *“Desarrollo de un modelo predictivo de diabetes mellitus con redes neuronales artificiales utilizando Python”*, con el objetivo de construir un modelo predictivo basado en redes neuronales. La

metodología fue experimental, empleando programación en Python y datos clínicos. Los resultados mostraron un desempeño adecuado del modelo en la predicción de la diabetes. Se concluye que las redes neuronales artificiales son herramientas eficaces para el apoyo diagnóstico.

Por otro lado, Torres M. y Palomino C. ⁽²²⁾ realizaron en Lima Metropolitana, en el año 2022, el estudio titulado “*Prevalencia de diabetes mellitus tipo 2 y factores asociados*”, con el objetivo de determinar la prevalencia de la diabetes tipo 2 y sus principales factores asociados en la población adulta. La metodología fue de tipo observacional y transversal, utilizando datos epidemiológicos y encuestas de salud. Los resultados evidenciaron una alta prevalencia de diabetes mellitus tipo 2, asociada principalmente a factores como edad, obesidad, sedentarismo y antecedentes familiares. Se concluye que la diabetes representa un problema de salud pública relevante en Lima Metropolitana, requiriendo estrategias preventivas y de detección temprana más efectivas.

Del mismo modo, Kumar V. y Sharma M. ⁽²³⁾ en la India del en el año 2022 desarrollaron el estudio denominado “*Comparative Analysis of Machine Learning Techniques for Diabetes Prediction*”, el objetivo fue comparar distintas técnicas de aprendizaje automático para la predicción de la diabetes. La metodología consistió en un enfoque comparativo cuantitativo, utilizando un conjunto de datos estandarizado y aplicando algoritmos como KNN, SVM, Random Forest y redes neuronales. Los resultados mostraron variaciones significativas en la precisión de los modelos, siendo Random Forest y SVM los que obtuvieron mejor rendimiento predictivo. Los autores concluyen que la selección adecuada del algoritmo es crucial para mejorar la precisión de los sistemas predictivos en salud y que los modelos de aprendizaje automático pueden apoyar eficazmente la detección temprana de la diabetes.

A su vez, en el Perú en el año 2021 en un artículo de investigación realizado por Mendoza et al. ⁽²⁴⁾, titulado “*Aplicación de modelos predictivos para la detección de enfermedades crónicas en el primer nivel de atención*”, tuvo como objetivo evaluar la utilidad de modelos predictivos en la detección temprana de enfermedades crónicas dentro del sistema de atención primaria. La metodología empleada fue de tipo cuantitativa, con un enfoque analítico, utilizando bases de datos clínicas provenientes de establecimientos de salud del primer nivel. Se aplicaron modelos estadísticos y de aprendizaje automático para identificar patrones de riesgo en los pacientes. Los resultados demostraron que los modelos

predictivos mejoraron significativamente la identificación temprana de pacientes con riesgo de enfermedades crónicas, optimizando la toma de decisiones clínicas. Como conclusión, los autores señalan que la incorporación de modelos predictivos en el primer nivel de atención puede fortalecer las estrategias de prevención y reducir la carga de enfermedades crónicas en el sistema de salud.

Además, León Velarde et al. ⁽²⁵⁾ realizaron en el Perú, en 2021, el estudio “*Acceso y uso de servicios de salud en zonas rurales del Perú: desafíos para la equidad*”, con el objetivo de analizar las barreras de acceso a los servicios de salud en zonas rurales. La metodología fue de tipo descriptivo y analítico, utilizando encuestas y registros administrativos. Los resultados evidenciaron limitaciones importantes en el acceso oportuno a servicios de diagnóstico y tratamiento de enfermedades crónicas, incluida la diabetes. Se concluye que las desigualdades en el acceso a la salud afectan directamente el diagnóstico temprano, por lo que es necesario fortalecer los servicios de atención primaria en zonas rurales.

A su vez, en el año 2020, Gómez F. y Sánchez M. ⁽²⁶⁾ desarrollaron el estudio en Lima, denominado “*Retos del diagnóstico temprano de la diabetes mellitus en zonas rurales peruanas*”, cuyo objetivo fue identificar las principales dificultades para el diagnóstico oportuno de la diabetes en contextos rurales. La metodología fue cualitativa y descriptiva, basada en revisión documental y análisis de información sanitaria. Los resultados señalaron limitaciones en infraestructura, personal capacitado y acceso a pruebas diagnósticas. Se concluye que estas barreras retrasan el diagnóstico de la diabetes, incrementando el riesgo de complicaciones.

En el presente estudio, la interrogante de investigación es: ¿cómo desarrollar un método de predicción para la Diabetes Mellitus con capacidades clínicas probabilísticas empleando un modelo de red neuronal artificial desarrollado en Python en la ciudad de Jaén-2023? La investigación se sustenta en la aplicación de la Inteligencia Artificial a la medicina, especialmente en el desarrollo de modelos de redes neuronales artificiales para identificar patrones no evidentes a simple vista en grandes volúmenes de datos clínicos, ya que, en la búsqueda del estado del arte, esta tecnología evidencia un rendimiento superior a las tecnologías existentes, alcanzando niveles de precisión que superan el 90 %.

La presente investigación tiene importancia social, pues radica en la calidad de vida de los habitantes de la provincia de Jaén. La diabetes mellitus es una enfermedad que avanza sin síntomas, a menudo se diagnostica cuando ya se han producido lesiones irreversibles. En este sentido, el modelo predictivo basado en inteligencia artificial puede proporcionar a los ciudadanos diagnósticos más tempranos, promoviendo estilos de vida saludables y aliviando el dolor familiar y personal que viene con las complicaciones.

Esta investigación fortalece el marco teórico existente sobre la aplicación de redes neuronales artificiales en el ámbito clínico, especialmente para la predicción de la Diabetes Mellitus. Al momento de analizar el comportamiento de variables metabólicas claves como la glucosa basal, la HbA1c y la prueba de tolerancia oral a la glucosa, las cuales mostraron correlaciones altas y a su vez consistentes. La investigación pretende el corroborar la relevancia de estos indicadores clínicos como predictores confiables, reforzando así los fundamentos expuestos en la literatura internacional. Además, el integrar enfoques de aprendizaje profundo con criterios de diagnósticos médicos permitirá ampliar la base conceptual sobre el uso de modelos de LSTM en el campo de la salud, logrando articular conocimientos biomédicos con capacidades computaciones avanzadas. De igual manera, esta investigación aspira a ampliar el sustento teórico que se tiene acerca de cómo las tecnologías predictivas pueden emplearse como herramientas de apoyo al diagnóstico clínico. Así, este aporte teórico abre la posibilidad de que futuras investigaciones adapten modelos neuronales a distintos contextos epidemiológicos, especialmente en zonas con recursos limitados como la provincia de Jaén.

Desde la perspectiva práctica, esta investigación aportará una herramienta tecnológica aplicable en el entorno real de los centros de salud de la provincia de Jaén. Este modelo predictivo basado en LSTM, entrenado con registros clínicos locales, buscará facilitar la identificación temprana de cuyos pacientes muestren riesgo de diabetes o prediabetes. Con ello, se espera optimizar el tiempo de atención, apoyar al personal de salud en la toma de decisiones y mejorar la capacidad diagnóstica en aquellos establecimientos con limitaciones de personal, equipamiento o recursos. Asimismo, el modelo podrá convertirse en insumo valioso para futuras integraciones en plataformas digitales de salud o sistemas de historias clínicas. Este aporte práctico tendrá potencial de mejorar la cobertura diagnóstica, fortaleciendo la prevención y el tamizaje temprano en el primer nivel de atención.

Como aporte metodológico, la investigación plantea una rigurosa aplicación del modelo CRISP-DM, una metodología ampliamente aceptada para proyectos de minería de datos. Considerando que cada etapa servirá como una guía estructurada que garantizará la calidad del proceso analítico. Además, la incorporación de herramientas modernas como Sckit-learn, TensorFlow, técnicas de estandarización, validación cruzada y métodos de interpretación como SHAP que buscará demostrar que es posible combinar métodos estadísticos tradicionales con técnicas avanzadas de aprendizaje profundo. Esta propuesta metodológica permitirá que otras investigaciones similares puedan replicar y mejorar el procedimiento, contribuyendo a la formación metodológica en salud digital dentro del contexto local.

Finalmente, el estudio aportará conocimiento científico al describir el perfil clínico y metabólico de una población representativa, los valores reales de glucosa, HbA1c, IMC y pruebas de tolerancia oral. Estos datos ayudarán a conocer mejor la situación de la diabetes y podrán servir como base para nuevas investigaciones, vinculada al uso de tecnologías predictivas para mejorar la vigilancia epidemiológica de enfermedades crónicas.

En este contexto y considerando la necesidad de mejorar la detección temprana de la Diabetes Mellitus en la provincia de Jaén, para la presente investigación se formuló como objetivo general: desarrollar un método de diagnóstico para la Diabetes Mellitus con capacidades clínicas probabilísticas empleando un modelo de red neuronal artificial desarrollado en Python en la ciudad de Jaén-2023. Este objetivo refleja la intención de aprovechar herramientas tecnológicas accesibles que permitan identificar riesgos de manera oportuna y a la vez apoyar el trabajo del personal de salud.

Asimismo, para orientar el desarrollo del presente estudio, se plantearon dos objetivos específicos, las cuales nos permitirán guiar cada etapa del proceso de investigación. El primero es: determinar los datos de diagnóstico clínicos relevantes sobre la Diabetes Mellitus que serán utilizados para el desarrollo del modelo de red neuronal artificial en Python, lo cual permitirá seleccionar las variables más importantes y con mayor valor predictivo. El segundo es: determinar la capacidad de la red neuronal artificial desarrollada en Python para predecir Diabetes Mellitus en la ciudad de Jaén-2023, evaluando su desempeño y por ende verificando si realmente puede convertirse en una herramienta útil para el diagnóstico temprano.

II. MATERIAL Y MÉTODOS

2.1. Ubicación

El presente estudio se llevó a cabo en la provincia y distrito de Jaén, región Cajamarca, ubicada en el norte del Perú. La base de datos sobre la diabetes mellitus, se obtuvo a partir de los reportes de laboratorio y las historias clínicas de pacientes que recibieron atención en el Hospital San Javier de Bellavista, institución ubicada en esta misma provincia y que atiende a una parte importante de la población local.

2.2. Población, muestra y muestreo

Población:

La población objeto de estudio estuvo conformada por todas las historias clínicas de pacientes jóvenes y adultos entre 18 y 70 años respectivamente, atendidos en el Hospital San Javier de Bellavista. Los pacientes han sido diagnosticados con Diabetes Mellitus o presentan múltiples factores de riesgo que nos indican una alta probabilidad de padecer esta enfermedad. La elección de estas historias clínicas se basa en la necesidad de obtener información clínica y de laboratorio que sea completa y de alta calidad, incluyendo métricas importantes como niveles de glucosa en ayunas, hemoglobina glicosilada, prueba de tolerancia a la glucosa, índice de masa corporal y mediciones de presión arterial.

Muestra:

Para este estudio la muestra fue por conveniencia y se recolectó información mediante la revisión de 415 historias clínicas de pacientes jóvenes y adultos entre 18 y 70 años además de los resultados de laboratorio, los cuales fueron atendidos en el Hospital San Javier de Bellavista.

Criterios de inclusión

- Historias clínicas de pacientes entre 18 y 70 años con cualquier condición de salud, atendidas en el Hospital San Javier de Bellavista.
- Historias clínicas de pacientes con un registro médico completo, que incluya información sociodemográfica, clínica y bioquímica, como IMC, presión arterial, niveles de glucosa en ayunas, HbA1c, prueba de tolerancia oral a la glucosa (OGTT).

- Historias clínicas que incluyan una confirmación médica de ausencia o presencia de prediabetes o diabetes mellitus.
- Historias clínicas con autorización registrada para el uso de la información con fines de investigación y estadísticos, asegurando el anonimato y la confidencialidad.

Criterios de exclusión

- Historias clínicas correspondientes a pacientes menores de 18 años o mayores de 70 años.
- Historias clínicas con registro médico incompleto o que carecen de información en alguna de las variables relevantes para el estudio.
- Historias clínicas de pacientes que presentan enfermedades crónicas severas concurrentes, que podrían afectar la evaluación de la diabetes.
- Historias clínicas con diagnóstico de diabetes incorrecto o incierto, o cuyos resultados no se han verificado mediante pruebas convencionales.
- Historias clínicas duplicadas o con información redundante en la base de datos.

Muestreo:

Se empleó un muestreo de conveniencia no probabilístico, seleccionando las historias clínicas de los pacientes que cumplieran con los criterios de inclusión previamente establecidos, en lugar de un muestreo aleatorio, lo que garantizó la pertinencia y coherencia de los datos. También se definieron y aplicaron criterios de exclusión para garantizar la pertinencia de los datos analizados ⁽²⁹⁾.

Variables de estudio

Las variables identificadas en el presente estudio son:

- **Variable dependiente:** Presencia de Diabetes Mellitus.
- **Variable independiente:** Factores epidemiológicos.

Operacionalización de variables: ver anexo 01

2.3. Método, técnica e instrumento de recolección de datos

Método:

Inicialmente, se envió un documento formal dirigido al responsable de la dirección del Hospital San Javier de Bellavista, mediante el cual se solicitó la autorización para acceder a las historias clínicas y a los resultados de laboratorio. Una vez autorizadas, se realizó la coordinación correspondiente con el área de laboratorio para la obtención de los resultados bioquímicos; posteriormente, se gestionó el acceso con el área de archivos para recopilar la información contenida en las historias clínicas. Luego de reunir la totalidad de los datos, se efectuó un proceso de limpieza y depuración de datos, eliminando registros duplicados o repetitivos. Los datos seleccionados fueron entregados al programador para su procesamiento y para el desarrollo del modelo de red neuronal artificial tipo LSTM. Los resultados obtenidos fueron analizados y sistematizados, lo que permitió la elaboración del informe final de la investigación.

Técnica e instrumento de recolección de datos:

La técnica empleada en la presente investigación fue la observación directa, aplicada durante la revisión ordenada de las historias clínicas y de los registros de laboratorio del Hospital San Javier de Bellavista. A través de esta técnica se identificó y verificó la información relevante para el estudio, asegurando que los datos recopilados cumplieran con los criterios establecidos, sin realizar intervención alguna sobre los pacientes.

El instrumento utilizado fue una ficha de registro de datos (*ad hoc*), diseñada específicamente para la extracción sistemática de la información contenida en las historias clínicas y en los resultados de laboratorio. Esta ficha permitió recopilar de manera ordenada los datos clínicos y bioquímicos necesarios para el estudio, asegurando la consistencia y calidad de los datos que posteriormente fueron procesados en el modelo de red neuronal artificial. El documento fue validado mediante el juicio de expertos, contando con la evaluación de tres especialistas, se adjunta en el anexo 3.

2.4. Tipo, diseño y método de investigación

La presente investigación según su finalidad es de tipo aplicada; según su nivel, es explicativa; según su diseño, es no experimental de corte transversal, de enfoque cuantitativo y de método de investigación hipotético-deductivo.

Aplicada: Es de tipo aplicada, ya que tiene como finalidad resolver problemáticas concretas de la sociedad mediante el uso de conocimiento derivados de la investigación básica, ya que al desarrollar un modelo de red neuronal artificial en lenguaje de Python con el objetivo de predecir la presencia de Diabetes Mellitus, este modelo emplea variables clínicas y demográficas, donde su aplicación busca mejorar la detección temprana de esta enfermedad, lo cual contribuye así a la mejora del sistema de salud y bienestar de los pacientes de la provincia de Jaén ⁽²⁷⁾.

El ser de tipo aplicada nos permite aspirar a ofrecer una respuesta efectiva y a su vez concreta, ésta responde a un desafío de salud pública a través de la utilización de nuevas tecnologías. Por ende, la investigación aplicada tiene como objetivo el hacer uso del conocimiento científico para fines prácticos, donde se aborda los problemas de manera específica, en este caso el estudio busca contribuir con la identificación temprana de la Diabetes Mellitus, siendo de apoyo a un diagnóstico oportuno y a su vez una mejora en la atención ⁽²⁸⁾.

Explicativa: El nivel de la investigación es explicativo, ya que no se limita a examinar la presencia de Diabetes Mellitus en pacientes de la provincia de Jaén, sino que también pretende identificar y entender cómo diferentes variables demográficas y clínicas (como edad, sexo, índice de masa corporal, presión arterial promedio, Glucemia en ayunas, HbA1c, prueba de tolerancia oral a la glucosa) pueden estar relacionadas con esta patología. Permite descubrir las causas que originan un suceso, lo que a su vez posibilita establecer las relaciones de causa-efecto entre las variables implicadas ⁽²⁹⁾.

No experimental: La investigación se basa en un diseño no experimental, ya que no se manipulan directamente las variables, los datos previamente recolectados son analizados tal como se presentan. No hay intervención en las condiciones que presentan los pacientes, sino que son de uso de estudio en cuanto a sus características clínicas y demográficas para identificar patrones asociados a esta patología. Además, es de corte transversal, porque la información es recolectada y analizada en un momento determinado del tiempo, lo que nos permite obtener un panorama del fenómeno en ese punto específico. Esta característica es importante porque nos permite evaluar el comportamiento de múltiples variables en un momento dado, sin que haya necesidad de realizar un seguimiento a los pacientes a lo largo del tiempo ⁽²⁹⁾.

Cuantitativo: Esta investigación adopta un enfoque cuantitativo y utiliza variables numéricas para desarrollar un modelo matemático que permita predecir la enfermedad. Este tipo de enfoque permite establecer relaciones causales o predictivas a través de procedimientos estadísticos, algoritmos de aprendizaje automático (Machine Learning) y pruebas de precisión ⁽³⁰⁾.

Hipotético-deductivo: La investigación adopta un método hipotético-deductivo, ya que partió de ideas generales sobre el uso de redes neuronales en salud aplicadas a la predicción de la Diabetes Mellitus en el Hospital San Javier de Bellavista de la Provincia de Jaén. Asimismo, a partir de los conocimientos previos, se lograron formular las hipótesis, que posteriormente a ello fueron comprobadas con la información obtenida en el estudio, es decir, la base de datos. Según la metodológica, este método va desde lo general hacia lo particular, validando así la aplicabilidad en contextos concretos y permitiendo encontrar relación con las variables establecidas para la investigación ⁽²⁸⁾.

De este modo, el desarrollo de un modelo de red neuronal artificial tiene como finalidad el predecir un resultado concreto (presencia de diabetes) a partir de múltiples variables independientes, con esto se muestra un enfoque explicativo. Aunque el diseño no incluye manipulación de las variables de forma directa, el modelo realiza el análisis profundo con una base de datos ya existente, lo cual nos permite establecer relaciones explicativas, en este caso sobre la diabetes ⁽²⁹⁾.

2.5. Procedimiento para la recolección de datos

Una vez obtenida la autorización por parte del director del Hospital San Javier de Bellavista, para la recopilación de datos, se firmó el acta de confidencialidad correspondiente. Posteriormente, se coordinó con el personal designado de las áreas involucradas para informales los objetivos del estudio y el procedimiento a seguir; estableciendo criterios de acceso, revisión y registro de la información. Asimismo, se acordó que los datos serían tratados con estricta confidencialidad mediante un sistema de codificación numérica, con fines exclusivamente investigativos, garantizando el anonimato de los titulares de las historias clínicas y resguardando su identidad.

Luego de dichas coordinaciones, el encargado del área de laboratorio nos facilitó el acceso a los registros que contenían las variables del estudio: (glucosa basal, hemoglobina glicosilada, tolerancia oral a la glucosa, N° historia clínica). Se organizó el proceso con el

fin de asegurar el uso adecuado de ficha *ad hoc* como herramienta recolectora de datos y la auditoría de las historias clínicas.

De manera complementaria, se accedió a la base de datos del área de admisión, de donde se obtuvieron las variables sociodemográficas y clínicas, tales como edad, género, peso, talla, presión arterial; además de la verificación de los números de historias clínicas. A partir de esta información, se realizó el filtrado de los registros que cumplieran con los criterios de inclusión establecidos, depurando aquellos que no cumplieran con dichos criterios.

Durante todo este procedimiento se cumplieron los lineamientos vigentes relacionados con el manejo de datos sensibles. La información se obtuvo mediante análisis documental de historias clínicas y resultados de laboratorio, registrándose de forma ordenada y verificable para asegurar precisión y validez.

En relación con las fases para la construcción del modelo de red neuronal artificial, este proceso se desarrolló por etapas: primero se comprendió el problema y su finalidad sanitaria (entendimiento del negocio); luego se identificaron y revisaron las variables disponibles y su calidad (entendimiento de los datos); a continuación se consolidó, depuró y estructuró la base de datos (preparación de los datos); seguidamente se construyó el modelo en Python ajustando su arquitectura y parámetros (modelamiento); después se midió su desempeño con métricas de clasificación y validación para confirmar su capacidad de generalización (evaluación); y, finalmente, se planteó su posible integración como herramienta de apoyo en entornos digitales de salud, incorporando técnicas de interpretabilidad para facilitar su comprensión por profesionales (implantación).

Con toda la información recolectada, se procedió a la consolidación en una base digital, utilizando Microsoft Excel® como apoyo para organizar los registros y facilitar la detección de inconsistencias. En esta fase se ejecutaron acciones clave de depuración y estandarización, tales como: la eliminación de registros duplicados o repetidos, revisión y corrección de inconsistencias entre variables, verificación de rangos razonables en variables numéricas, estandarización del formato de datos (por ejemplo, unidades y tipos de dato), control de valores faltantes para definir su tratamiento antes del modelamiento.

Una vez validada la base depurada, el conjunto de datos se dividió para el proceso de entrenamiento y evaluación: 20% para validación y 80% para entrenamiento. Esta partición

permitió entrenar el modelo con la mayor cantidad de información posible y a la vez, medir su desempeño con datos no vistos durante el entrenamiento.

Modelamiento: Al llegar a esta etapa, una vez que logramos obtener los datos debidamente organizados, procedimos a importar dicha base al entorno de programación en Python. Elegimos este lenguaje de programación que nos permitió la creación del modelo de red neuronal, para ello utilizamos bibliotecas especializadas para aprendizaje automático y redes neuronales, como Scikit-learn y TensorFlow, las cuales son herramientas complementarias para preprocesamiento y evaluación.

En esta etapa, nos centramos en definir la arquitectura del modelo. Para ello, establecimos tres niveles fundamentales: la capa de entrada, configurada según el número de variables clínicas recolectadas; las capas ocultas, donde determinamos la estructura y número de neuronas necesarias para el aprendizaje de patrones; y, finalmente, la capa de salida, encargada de arrojar la predicción sobre la presencia o ausencia de diabetes mellitus, de acuerdo con el enfoque planteado en nuestro estudio.

A lo largo del desarrollo, ejecutamos diversas pruebas de configuración con el objetivo de optimizar el rendimiento. Esto implicó ajustar minuciosamente parámetros como la tasa de aprendizaje, el número de épocas, el tamaño de lote (batch size) y las funciones de activación. Estos ajustes técnicos permitieron mejorar la capacidad del modelo para reconocer patrones clínicos relevantes sin perder estabilidad en el entrenamiento.

Evaluación del modelo: Para garantizar que el modelo de red neuronal desarrollado no solo tuviera un buen desempeño con los datos de entrenamiento, sino que también pudiera generalizar adecuadamente ante nuevos casos en contextos reales, para ello se aplicaron estrategias orientadas a controlar el sobreajuste y fortalecer la confiabilidad de los resultados, incluyendo validación con el subconjunto reservado y técnicas de regularización cuando correspondió.

El desempeño se midió mediante métricas habituales en clasificación, tales como: precisión (accuracy), sensibilidad (recall), especificidad y curva ROC para observar el comportamiento del modelo en diferentes umbrales. Además, se revisaron indicadores de error (como MSE) y medidas complementarias que permitieran identificar oportunidades de mejora, evaluando de manera crítica si el modelo se ajustaba al objetivo planteado y si mantenía estabilidad al aplicarse sobre datos de validación.

Implantación del Modelo: Luego de haber realizado la evaluación, se planteó como proyección la integración del modelo en entornos digitales de apoyo clínico, con el fin de contribuir a la automatización de alertas y a la detección temprana de la diabetes mellitus.

Para una implementación efectiva se requiere coordinación con instituciones de salud, definición de flujos de uso, capacitación del personal y validaciones adicionales en escenarios reales.

Considerando que uno de los principales desafíos al momento de trabajar con redes neuronales artificiales es la limitada interpretabilidad de sus resultados, incluso cuando se muestran altos niveles de precisión, por ello para resolver esta dificultad y facilitar su comprensión a los profesionales de la salud, se propuso complementar el modelo con técnicas explicativas como SHAP (Shapley Additive exPlanations); que permiten visualizar el aporte de cada variable en la predicción final. Esto facilita que el personal de salud comprenda el comportamiento del modelo, mejore la confianza en los resultados y use la herramienta como apoyo, sin reemplazar el juicio clínico.

2.6. Análisis estadístico de datos

Una vez recolectada la información en una ficha de datos ad hoc (Anexo 3); se procedió a pasar toda la información al programa Microsoft Office Excel® 2021 y el análisis de datos se realizó con el lenguaje de programación Python.

Se desarrolló los modelos de redes neuronales artificiales, de forma aleatoria se eligió el 80 % de los datos para entrenamiento y el 20 % para prueba y obtención del margen de error. Luego de crear los modelos, se validó su confiabilidad, para ello se utilizaron indicadores como el error cuadrático medio y el coeficiente de determinación (r^2).

2.7. Consideraciones éticas

Este estudio se llevó a cabo con la autorización otorgada por el encargado del Centro de Salud Morro Solar y del Hospital San Javier De Bellavista (ver Anexo 2), lo cual permitirá la aplicación del instrumento y la revisión de las historias clínicas. Asimismo, es importante destacar que los datos serán tratados de forma confidencial, utilizando un sistema de registro numérico exclusivamente con fines de investigación, preservando y resguardando la identidad de los pacientes. Se cumplirá todas las normativas y regulaciones aplicables respecto a la privacidad y uso de datos.

III. RESULTADOS

Tabla 1. Datos de diagnóstico clínicos relevantes sobre diabetes mellitus

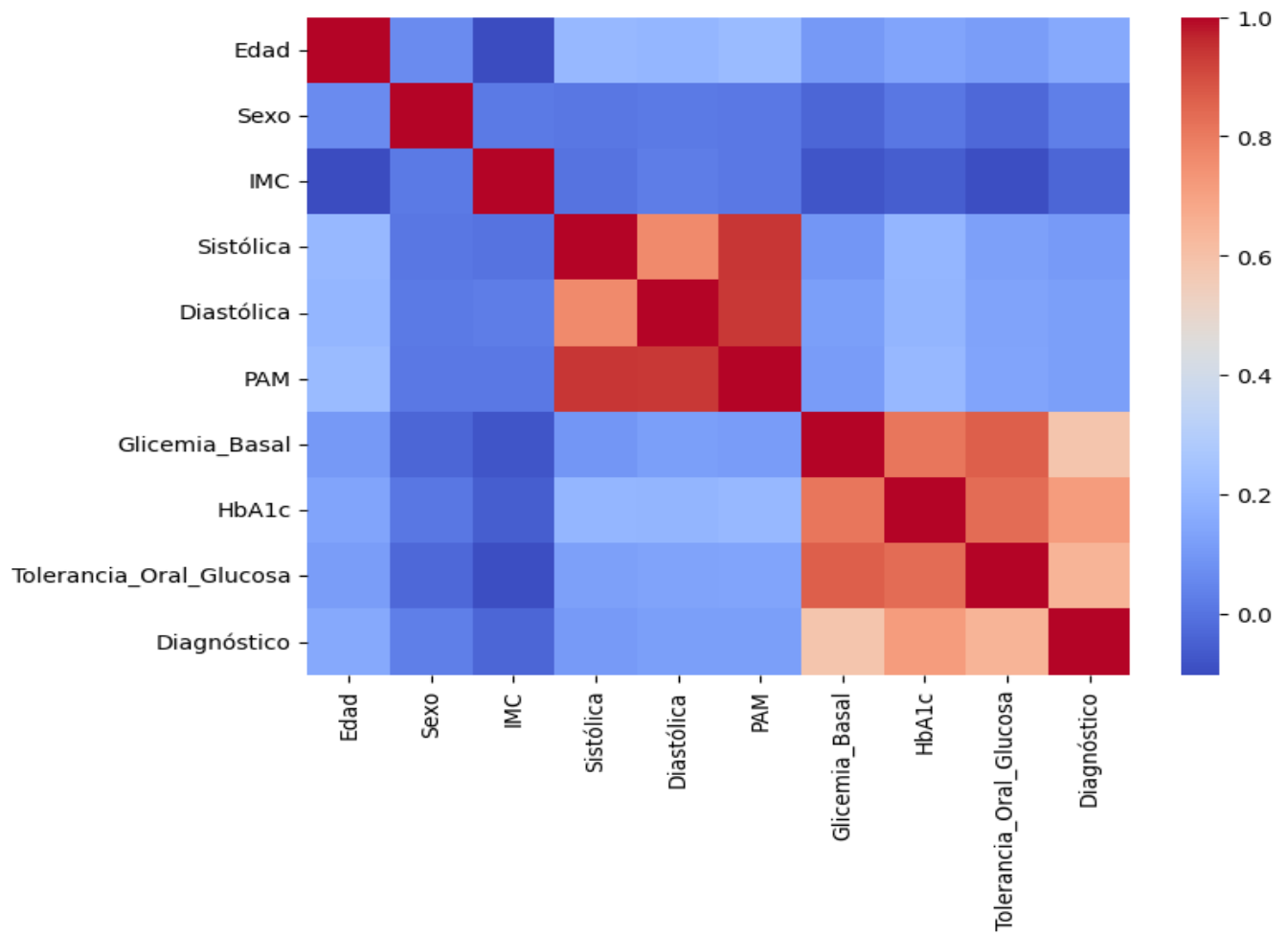
	Edad	genero	IMC	Sistólica	Diastólica	PAM	Glicemia_Basal	HbA1c	Tolerancia_Oral_Glucosa	Diagnóstico
count	415	415	415	415	415	415	415	415	415	415
mean	52.71	0.383	28.38	118.13	73.28	88.23	131.75	6.47	164.65	0.91
std	11.11	0.486	4.19	23.49	11.29	14.42	60.32	1.85	77.60	0.80
min	18	0	16.9	70	50	66.66	78	4	61	0
25%	45	0	25.75	110	65	80	100.09	5.5	118	0
50%	55	0	27.7	113	70	86.66	109	5.8	137	1
75%	61	1	30.45	120	80	93.33	133.95	6.65	184	2
max	70	1	46.4	254	132	172.66	470.3	15.9	628	2

Interpretación: Se llevó a cabo un análisis estadístico descriptivo de las variables clínicas consideradas relevantes para el diagnóstico de diabetes mellitus, a partir de 415 registros válidos de historias clínicas de pacientes adultos. La edad promedio de la muestra fue de 52.7 años, con un rango de 18 a 70 años, lo que permite una adecuada representación amplia de la población susceptible a desarrollar diabetes mellitus. La variable sexo se codificó de forma binaria (0 = femenino, 1 = masculino), obteniéndose una media de 0.38, lo que indica una mayor presencia de mujeres en la muestra. El índice de masa corporal (IMC) presentó un valor promedio de 28.38 kg/m², indicando una condición de sobrepeso, con valores máximos que llegaron hasta 46.4 kg/m², lo cual representa un factor de riesgo relevante ligado a la obesidad severa y al desarrollo de la enfermedad.

Los valores promedio de presión arterial fueron: sistólica 118.13 mmHg, diastólica 73.28 mmHg y presión arterial media 88.23 mmHg, todos dentro de rangos referenciales, aunque se observaron casos con cifras elevadas compatibles con hipertensión. En cuanto a los indicadores metabólicos, la glucemia en ayunas presentó una media de 131.75 mg/dL, valor superior al umbral diagnóstico de diabetes mellitus. De manera similar, la hemoglobina glicosilada (HbA1c) alcanzó un promedio de 6.47 % cercano al punto de corte diagnóstico de 6.5 %, mientras que la prueba de tolerancia oral a la glucosa mostró un valor 164.65 mg/dL también superior de los rangos referenciales. Estos hallazgos evidencian la presencia significativa de alteraciones metabólicas en la población evaluada.

Finalmente, la variable dependiente (diagnóstico clínico), fue codificada en tres categorías: 0 (sin diabetes), 1 (prediabetes) y 2 (diabetes confirmada), obteniéndose una media de 0.91, lo que confirma una elevada proporción de pacientes con riesgo metabólico o con diagnóstico confirmado. En conjunto, los datos presentados en la tabla 1 permiten identificar los principales indicadores clínicos relevantes para el diagnóstico de diabetes mellitus y justifican su selección como variables de entrada para el modelo de red neuronal artificial.

Figura 1. Mapa de calor para determinar la relevancia de las variables



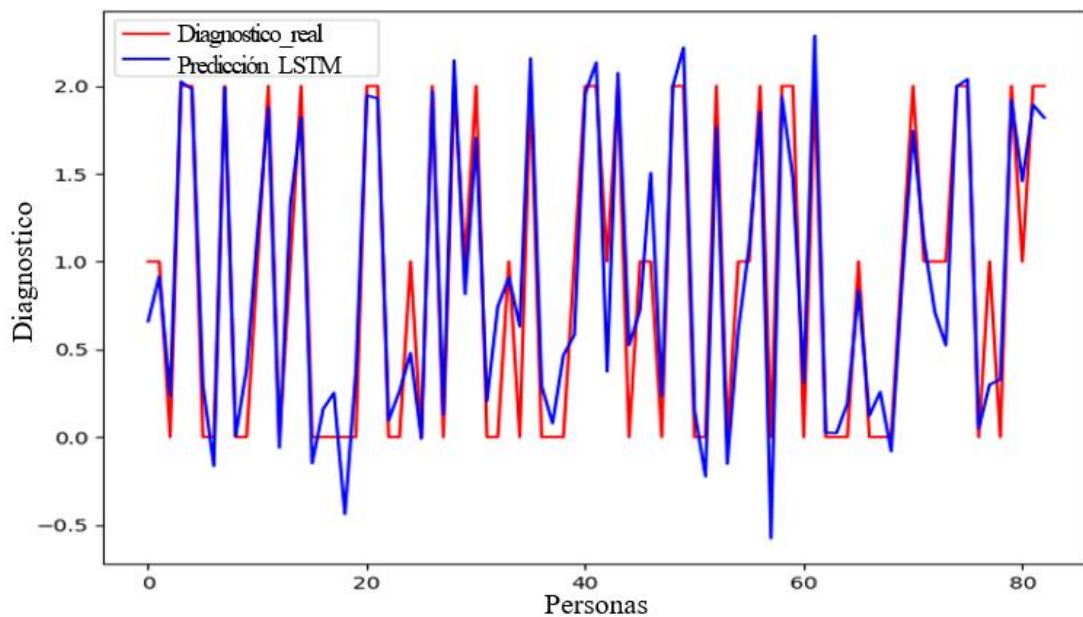
Interpretación: La matriz de correlación obtenida permite evaluar la intensidad y dirección de las relaciones lineales entre las variables incluidas en el modelo de predicción de Diabetes Mellitus propuesto para la ciudad de Jaén. Conforme a la interpretación estadística del coeficiente de correlación de Pearson, los valores próximos a ± 1 indican asociaciones lineales fuertes, mientras que aquellos cercanos a 0 evidencian independencia lineal entre las variables. El análisis revela una correlación significativa entre el diagnóstico de diabetes y tres variables clínicas fundamentales: la glicemia basal ($r = 0.586$), la hemoglobina glicosilada HbA1c ($r = 0.718$) y la tolerancia oral a la glucosa ($r = 0.643$). Esta fuerte asociación no solo valida clínicamente el enfoque del modelo, sino que también respalda su coherencia con los criterios diagnósticos establecidos por organismos internacionales como la ADA (American Diabetes Association), consolidando su valor como predictores claves en sistemas inteligentes de clasificación.

En paralelo, se identifica una correlación altamente positiva entre la presión sistólica, la presión diastólica y la presión arterial media (PAM), siendo esta última una función ponderada de las dos primeras ($r = 0.942$ con sistólica y $r = 0.937$ con diastólica), lo cual corrobora la consistencia interna de los datos fisiológicos. Aunque estas variables muestran una correlación moderada con la hemoglobina glicosilada y otras métricas glucémicas, su aporte directo al diagnóstico es menos determinante ($r \approx 0.12$), lo cual sugiere su inclusión como posibles moduladores secundarios en modelos multivariados.

Por el contrario, las variables edad ($r = 0.155$), género ($r = 0.027$) y el índice de masa corporal (IMC) ($r = -0.036$) exhiben correlaciones débiles con el diagnóstico de diabetes en este conjunto de datos. Esto sugiere que, en el contexto poblacional específico de Jaén para el año 2023, estas variables no poseen un peso predictivo significativo de forma independiente. Sin embargo, podrían aportar valor en modelos no lineales o como variables latentes cuando se integran con otros factores de riesgo.

El análisis de correlación constituye una etapa preliminar crítica para la optimización del modelo de red neuronal artificial, permitiendo priorizar las variables más influyentes y reducir la dimensionalidad sin comprometer la precisión predictiva. Además, ofrece una base empírica sólida para el ajuste de pesos sinápticos en la arquitectura neuronal, maximizando la eficiencia del aprendizaje supervisado en la detección de patrones relacionados con la Diabetes Mellitus tipo 2.

Figura 2. Validación gráfica del modelo de red neuronal artificial



Interpretación: La Figura 2 ilustra la validación del método de diagnóstico propuesto, basado en una red neuronal artificial LSTM, al comparar los valores reales de diagnóstico con las predicciones generadas por el modelo. En el gráfico se observa que el sistema logra reproducir, en términos generales, el comportamiento del diagnóstico clínico real, evidenciando la capacidad del modelo para identificar tendencias asociadas al estado metabólico de los pacientes.

No obstante, también se identifican diferencias entre los valores reales y las predicciones, las cuales se manifiestan en variaciones puntuales y en la presencia de valores fuera del rango esperado. Estas discrepancias reflejan errores de predicción atribuibles a la sensibilidad del modelo frente a la variabilidad de los datos clínicos y al ruido presente en la base de datos.

En este sentido, dichas variaciones se presentan como parte del proceso de validación del método, ya que permiten evidenciar la sensibilidad del modelo ante cambios en los datos clínicos y su comportamiento frente al ruido presente en la base de datos. Por ello, la figura muestra que, aunque existan diferencias puntuales, las predicciones generadas mantienen relación con el diagnóstico clínico real en términos generales.

A pesar de estas variaciones, el método desarrollado permite diferenciar entre los estados de ausencia de diabetes, prediabetes y diabetes, lo que evidencia su capacidad para generar diagnósticos con enfoque probabilístico. En consecuencia, los resultados mostrados en la figura indican que el modelo de red neuronal artificial constituye una base válida para el desarrollo de un método de diagnóstico de la Diabetes Mellitus con capacidades clínicas probabilísticas, susceptible de ser optimizado mediante ajustes adicionales de normalización y parámetros de entrenamiento.

IV. DISCUSIÓN

Según el objetivo específico 01: *Determinar los datos de diagnóstico clínicos relevantes sobre diabetes mellitus a utilizar para el desarrollo del modelo de red neuronal artificial en Python en la ciudad de Jaén-2023*. En nuestra investigación el análisis clínico realizado en el Hospital San Javier de Bellavista en Jaén se centra en la aplicación de criterios diagnósticos internacionales para la DM2, evaluando parámetros objetivos como glicemia basal, hemoglobina glicosilada (HbA1c) y prueba de tolerancia oral a la glucosa. Este estudio confirma la validez de dichos indicadores para discriminar entre pacientes normo glucémicos y aquellos con diabetes o prediabetes no controlada. La precisión en la identificación de valores alterados (glicemia ≥ 128 mg/dL, HbA1c ≥ 6.3 %, tolerancia oral ≥ 180 mg/dL) evidencia la importancia del diagnóstico clínico temprano y del seguimiento metabólico en la práctica hospitalaria. El estudio realizado por Torres y Palomino (2022) ⁽²²⁾ en Lima Metropolitana aporta una visión epidemiológica amplia sobre la prevalencia de la diabetes mellitus tipo 2 (DM2) y sus principales factores asociados en la población adulta. permitió identificar que la edad avanzada, la obesidad, el sedentarismo y los antecedentes familiares son los determinantes más relevantes para el desarrollo de la enfermedad. A partir de estos hallazgos, podemos reafirmar que la Diabetes Mellitus tipo 2, se ha consolidado como un problema de salud pública de gran magnitud. Frente a esta realidad, urgen estrategias preventivas y de detección temprana que sí sean efectivas en los contextos urbanos. En conjunto, los datos respaldan la selección de las variables mencionadas como insumos fundamentales para la construcción de un modelo basado en redes neuronales, alineado con la evidencia científica contemporánea nacional e internacional, demostrando que la inteligencia artificial es un aliado para apoyar el diagnóstico temprano de Diabetes Mellitus tipo 2 en contextos de baja cobertura.

Según el objetivo específico 02: *Determinar la capacidad de la red neuronal artificial desarrollada en Python para predecir diabetes mellitus en la ciudad de Jaén-2023*. En cuanto a su efectividad, la red neuronal demostró una alta capacidad para predecir correctamente la diabetes mellitus tipo 2, al identificar con notable precisión las alteraciones en los patrones clínicos dentro de los registros de pacientes de Jaén. Además, el sistema logró distinguir tanto los casos severos como aquellos perfiles intermedios compatibles con prediabetes. Reflejando una sensibilidad adecuada y alineada con criterios clínicos internacionales. El principal aporte de este modelo radica en su capacidad de emular con

precisión los criterios médicos establecidos, incluso en un entorno con limitaciones tecnológicas y acceso restringido al diagnóstico temprano. La concordancia entre las predicciones del sistema y los valores clínicos reales valida su aplicabilidad como herramienta complementaria en el primer nivel de atención. Estos resultados coinciden con investigaciones previas en Latinoamérica, donde modelos similares alcanzaron entre 77 % y 91 % de precisión ⁽¹⁸⁾, ⁽²⁰⁾. Sin embargo, el valor diferencial de esta propuesta radica en haber sido entrenada con datos clínicos locales, lo que potencia su utilidad en el contexto específico de la región Cajamarca. En síntesis, el modelo no solo predice con efectividad, sino que representa una solución viable y contextualizada para fortalecer el diagnóstico temprano de la diabetes tipo 2 en poblaciones vulnerables.

Según el objetivo general: *desarrollar un método de diagnóstico para la Diabetes Mellitus con capacidades clínicas probabilísticas empleando un modelo de red neuronal artificial desarrollado en Python en la ciudad de Jaén-2023*. Los resultados obtenidos mediante el análisis de correlación de Pearson demuestran la viabilidad estadística y clínica del modelo de red neuronal artificial propuesto para el diagnóstico de Diabetes Mellitus en Jaén (2023). En particular, se identificaron fuertes asociaciones entre variables claves del metabolismo glucémico —glicemia basal ($r = 0.586$), HbA1c ($r = 0.718$) y tolerancia oral a la glucosa ($r = 0.643$)— y el diagnóstico final, lo que refuerza su valor predictivo dentro del sistema neuronal. Estos hallazgos son consistentes con lo reportado por Reinoso et al. Quienes validaron modelos IA en Ecuador con resultados robustos a pesar de conjuntos de datos limitados ⁽¹⁸⁾. La alta correlación entre variables hemodinámicas (presión sistólica, diastólica y PAM) confirma la calidad interna de los datos y su posible función como moduladores secundarios. Aunque presentan menor influencia directa en el diagnóstico, pueden enriquecer la sensibilidad del modelo cuando se integran con otros factores, tal como lo sugiere el trabajo de Huamán et al. Quienes utilizaron modelos híbridos en Lima alcanzando precisiones del 77 %. ⁽²⁰⁾. En contraste, variables como edad, género e IMC presentaron correlaciones débiles, lo que podría indicar una menor relevancia lineal en esta población específica. Sin embargo, dada la naturaleza no lineal de las redes neuronales artificiales, su utilidad no debe descartarse. Estudios como el de Pineda et al. que aplicaron redes en Python en contextos similares, resaltan cómo estas variables pueden adquirir peso significativo cuando se combinan con factores metabólicos y se entrenan bajo algoritmos supervisados ⁽²¹⁾. En síntesis, este análisis proporciona evidencia empírica sólida para

respaldar el objetivo general, al demostrar que es posible construir un modelo clínicamente relevante con capacidad predictiva eficiente mediante redes neuronales desarrolladas en Python. El uso de las variables altamente correlacionadas permite optimizar el diseño del modelo, reducir la complejidad computacional y mejorar la precisión, como han evidenciado estudios internacionales recientes en Corea e India ⁽¹⁶⁾ ⁽²³⁾.

V. CONCLUSIONES Y RECOMENDACIONES

5.1. Conclusiones:

El análisis estadístico demostró que el valor medio de la glicemia basal (131.75 mg/dL) y de la HbA1c (6.47 %) dejaba en evidencia una alta proporción de casos prediabéticos o diabéticos. Este hallazgo reafirmamos su utilidad como indicadores clave en modelos predictivos.

Asimismo, la red neuronal artificial desarrollada en Python logró predecir diabetes mellitus con un sólido nivel de confianza del 87 % en su rendimiento global. Este porcentaje, respaldado por una buena precisión predictiva, un error mínimo y una adecuada generalización, consolida su viabilidad y aplicabilidad clínica. Por lo tanto, el sistema propuesto se perfila como una herramienta sumamente efectiva y adaptable a las necesidades operativas del primer nivel de atención en salud.

El método propuesto, basado en una red neuronal artificial tipo LSTM, resultó técnicamente viable y clínicamente coherente para la identificación de patrones asociados al diagnóstico de la enfermedad. Los resultados evidencian que el modelo es capaz de procesar variables clínicas relevantes, permitiendo diferenciar entre pacientes sin diabetes, con prediabetes y con diabetes.

5.2. Recomendaciones:

Al director de la RIS Jaen, gestores de innovación, tomadores de decisiones y responsables de los establecimientos de salud de la provincia de Jaén considerar la incorporación progresiva de los métodos de diagnósticos basados en inteligencia artificial, como el modelo de red neuronal desarrollado en la presente investigación, como un apoyo complementario al diagnóstico clínico tradicional para fortalecer a detección temprana de la Diabetes Mellitus y optimizar la toma de decisiones oportunas. Asimismo, se sugiere que estas acciones se articulen con estrategias de promoción y prevención dirigidas a la población adulta de Jaén, especialmente a las personas con factores de riesgo, promoviendo la evaluación periódica de indicadores clínicos relevantes, a fin de favorecer intervenciones tempranas orientadas al control metabólico y a la reducción de complicaciones asociadas.

Al director del Hospital San Javier de Bellavista, al personal de salud y a las instituciones prestadoras de servicios médicos, especialmente aquellos que manejan historias clínicas electrónicas, fortalecer la recopilación sistemática y estandarizada de variables clínicas clave como glicemia basal, hemoglobina glicosilada (HbA1c), prueba de tolerancia oral a la glucosa, índice de masa corporal y presión arterial. Esta recomendación responde al primer objetivo específico, ya que la calidad y consistencia de estos datos resultan determinantes para el correcto funcionamiento de modelos predictivos basados en inteligencia artificial y para mejorar la confiabilidad del diagnóstico clínico de la Diabetes Mellitus.

Al coordinador de la escuela profesional de Tecnología Médica de la Universidad Nacional de Jaén, a los futuros investigadores, profesionales en inteligencia artificial aplicada a la salud continuar evaluando y optimizando modelos de redes neuronales artificiales, particularmente arquitecturas LSTM desarrolladas en Python, mediante el uso de conjuntos de datos más amplios y procesos de validación adicionales. Plantear esta recomendación, responde directamente a nuestro segundo objetivo específico, puesto que nos permitiría evaluar con mayor precisión la capacidad predictiva del modelo y consolidar su utilidad como herramienta de apoyo, para la detección temprana de la Diabetes Mellitus en la ciudad de Jaén.

VI. REFERENCIAS BIBLIOGRAFICAS

1. Magliano DJ, Boyko EJ. Atlas de Diabetes de la Federación Internacional de Diabetes - National Center for Biotechnology Information (NCBI) – Bookshelf. [Online].; 2024 [cited 2025 Julio 13. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK581934/>.
2. Aguirre Ascona Y, Saboya Rios N. Métodos de aprendizaje supervisado para la predicción de diabetes: una revisión sistemática de la literatura - Repositorio UpeU. [Online]. Lima, Perú: Repositorio Institucional de la Universidad Peruana Unión; 2024 [cited 2025 Julio 13. Available from: <https://repositorio.upeu.edu.pe/items/6b713fbd-0ce8-4719-be7b-08b46cf9e6c1>.
3. Anwar F, Qurat-Ul-Ain, Ejaz MY, Mosavi A. Análisis comparativo del diagnóstico de la diabetes mellitus mediante diferentes enfoques: una encuesta – Social Science Research Network. [Online].: Elsevier B.B.; 2021 [cited 2025 Julio 13. Available from: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3746487.
4. Sharma A, Prasad KDV, Chakrasali SV, Gowda D, Kumar C, Chatuverdi A, et al. Sistema de atención médica basado en visión artificial para la identificación de diabetes y sus tipos mediante IA. [Online].: Elsevier B.V.; 2023 [cited 2025 Julio 13. Available from: <https://www.sciencedirect.com/science/article/pii/S2665917423000879>.
5. Ejyiyi CJ, Qin Z, Amos J, Ejyiyi MB, Nnani A, Ejyiyi TU, et al. Un modelo de diagnóstico predictivo robusto para la diabetes mellitus utilizando algoritmos de aprendizaje automático incorporados por Shapley. [Online].: Elsevier B.V.; 2023 [cited 2025 Julio 13. Available from: <https://doi.org/10.1016/j.health.2023.100166>.
6. Lee JE, Jeon HJ, Lee OJ, Lim HG. Diagnóstico de diabetes mellitus mediante ultrasonido de alta frecuencia y red neuronal convolucional. [Online].: Elsevier B.V.; 2024 [cited 2025 Julio 13. Available from: <https://doi.org/10.1016/j.ultras.2023.107167>.
7. Hoyos W, Ruiz R, Hoyos K. Implementación de mapas cognitivos difusos con algoritmos genéticos para predecir diabetes mellitus tipo 2. [Online]. Medellin, Colombia: Instituto Tecnológico Metropolitano (ITM); 2024 [cited 2025 Julio 13. Available from: <https://doi.org/10.22430/22565337.3061>.

8. Cuenca Andrade G, Guamán Quinche R, Álvarez Pineda F, Chamba Zaragocin W, Guamán Quinche JO. Estudio de algoritmos de inteligencia artificial más utilizados para el diagnóstico de diabetes mellitus tipo 2. [Online]. Puyo, Ecuador: Universidad Estatal Amazónica; 2023 [cited 2025 Julio 13. Available from: <https://doi.org/10.54753/cedamaz.v13i1.1804>.
9. García Guillen J, Baday Torrez MZ. Desarrollo de red neuronal artificial para el diagnóstico clínico y probabilístico de Diabetes Mellitus. [Online]. Tegucigalpa, Honduras: Repositorio Institucional UNITEC; 2023 [cited 2025 Julio 13. Available from: <https://repositorio.unitec.edu/server/api/core/bitstreams/ec1782c2-6d58-4b42-a286-3c660666cb0b/content>.
10. Villena Chávez J, Manrique H, Pretell E. Diabetes Mellitus en el Perú: Impacto sobre la salud. Recomendaciones para prevención y atención integral. [Online]. Lima, Perú: Academia Nacional de Medicina del Perú (ANM); 2023 [cited 2025 Julio 13. Available from: <https://anmperu.org.pe/sites/default/files/4.%20Rev%20Diabetes%20Mellitus%20ANM.pdf>.
11. Albitres Flores , Bernabé Ortiz. Precisión diagnóstica de los criterios ADA 2020 para diabetes no diagnosticada en una población peruana. [Online].: John Wiley & Sons, Ltd; 2020 [cited 2025 Julio 13. Available from: <https://pubmed.ncbi.nlm.nih.gov/33002547>.
12. Seinfeld J, Sobrevilla A, Rosales ML, Ibáñez M, Ruiz D, Penny E, et al. Carga económica de la diabetes tipo 2 en el Perú: un estudio de costo de enfermedad que valora las diferencias de costos asociadas al nivel de control glucémico. [Online].: Elsevier Ireland Ltd.; 2024. Available from: <https://pubmed.ncbi.nlm.nih.gov/38584495>.
13. Villegas Cubas JE, Capuñay Uceda OE, Coronado Navarro AF, Delgado Chavarri AH, Osoreo Granda E. Sistemas de información para la Red Neuronal Convolutiva en la detección de diabetes usando imágenes de fondo de ojo. [Online]. Lima, Perú: Universidad Tecnológica del Perú (UTP); 2021 [cited 2025 Julio 13. Available from: <https://dialnet.unirioja.es/servlet/articulo?codigo=9826367>.
14. Romero Rojas M. Perfil epidemiológico de mujeres con riesgo de desarrollar diabetes mellitus tipo 2 en un centro de salud del distrito de Jaén, 2023. [Online]. Jaén, Perú: Repositorio Institucional – Universidad Nacional de Jaén; 2023 [cited 2025 Julio 13. Available from: <https://repositorio.unj.edu.pe/handle/UNJ/682>.
15. Cubas Núñez JL. Prevalencia de diabetes mellitus tipo 2 en pacientes adultos en el Centro de Salud Nueva Esperanza - Jaén, 2022. [Online]. Jaén, Perú: Repositorio Institucional – Universidad Nacional

de Jaén; 2024 [cited 2025 Julio 13. Available from: <https://repositorio.unj.edu.pe/handle/20.500.14689/866>.

16. Eun Lee J, Ju Jeon H, Joun Lee O, Gyun Lim H. Diagnóstico de diabetes mellitus mediante ultrasonido de alta frecuencia y red neuronal convolucional. [Online].; 2024 [cited 2025 Julio 18. Available from: <https://doi.org/10.1016/j.ultras.2023.107167>.
17. Zhang Y , Li H , Wang X , Chen M. Application of Machine Learning Algorithms in Diabetes Diagnosis Based on Clinical Data. [Online].; 2023 [cited 2025 Julio 15. Available from: [10.1016/j.jbi.2023.104285](https://doi.org/10.1016/j.jbi.2023.104285).
18. Reinoso M , Cedeño J , Delgado M. Evaluación de algoritmos de inteligencia artificial para el diagnóstico de la diabetes mellitus tipo 2. [Online].; 2023 [cited 2025 Julio 16. Available from: revistas.unesum.edu.ec.
19. Ahmad A , Zubair M , Khan S. A Deep Learning Model for Early Detection of Diabetes Using Neural Networks. [Online].; 2023 [cited 2025 Julio 15. Available from: [10.12669/pjms.39.2.7084](https://doi.org/10.12669/pjms.39.2.7084).
20. Huamán A , Rivera F , Callañaupa G. Implementación de modelos híbridos de predicción de diabetes con aprendizaje supervisado en WEKA. [Online].: Portal de revistas UNMSM; 2023 [cited 2025 Julio 16.
21. Pineda M , Ramos E , Fernández J. Desarrollo de un modelo predictivo de diabetes mellitus con redes neuronales artificiales utilizando Python. [Online].; 2023 [cited 2025 Julio 17.
22. Torres Mendoza B, Palomino Cabrera H. Prevalencia de diabetes mellitus tipo 2 y factores asociados en Lima Metropolitana. [Online].; 2022 [cited 2025 Julio 14. Available from: [10.35663/amp.2022.393.234](https://doi.org/10.35663/amp.2022.393.234).
23. Kumar V , Sharma M. Comparative Analysis of Machine Learning Techniques for Diabetes Prediction. [Online].; 2022 [cited 2025 Julio 16. Available from: [10.21817/indjce/2022/v13i1/221301278](https://doi.org/10.21817/indjce/2022/v13i1/221301278).
24. Mendoza Calderón J, Rojas Luna D, García Gonzáles C. Aplicación de modelos predictivos para la detección de enfermedades crónicas en el primer nivel de atención. [Online].; 2021 [cited 2025 Julio 14. Available from: [10.17843/rpmesp.2021.382.1046](https://doi.org/10.17843/rpmesp.2021.382.1046).

25. León Velarde F, Cárdenas R. , Huamán Espino L. Acceso y uso de servicios de salud en zonas rurales del Perú: desafíos para la equidad. [Online].; 2021 [cited 2025 Julio 14. Available from: [10.17843/rpmesp.2021.384.9043](https://doi.org/10.17843/rpmesp.2021.384.9043).
26. Gómez Pérez F, Sánchez M.. Retos del diagnóstico temprano de la diabetes mellitus en zonas rurales peruanas. [Online].; 2020 [cited 2025 Julio 14. Available from: [10.20453/rmh.v31i1.3742](https://doi.org/10.20453/rmh.v31i1.3742).
27. Ñaupas H, Mejía E, Novoa E, Villagómez A. Metodología Investigación Científica y Asesoramiento Tesis – Una propuesta didáctica para aprender a investigar y elaborar la tesis. [Online]. Lima: CEPREDIM; 2011 [cited 2025 Julio 13. Available from: http://www.biblioteca.cij.gob.mx/Archivos/Materiales_de_consulta/Drogas_de_Abuso/Articulos/MetodologiaInvestigacionNaupas.pdf.
28. Hernández R, Fernández C, Baptista M del P. Metodología de la investigación. [Online]. México, Bogotá, Buenos Aires, Caracas, Guatemala, Madrid, Nueva York, San Juan, Santiago, Sao Paulo, Auckland, Londres, Milán, Montreal, Nueva Delhi, San Francisco, Singapur, St. Louis, Sídney, Toronto: McGraw-Hill / Interamericana Editores S.A. de C.V.; 2014 [cited 2025 Julio 13. Available from: <https://www.esup.edu.pe/wp-content/uploads/2020/12/2.%20Hernandez,%20Fernandez%20y%20Baptista-metodolog%C3%ADa%20Investigacion%20Cientifica%206ta%20ed.pdf>.
29. Arias JL, Covinos M. Diseño y metodología de la investigación. [Online]. Arequipa: Enfoques Consulting EIRL; 2021 [cited 2025 Julio 13. Available from: https://gc.scalahed.com/recursos/files/r161r/w26022w/Arias_S2.pdf.
30. Creswell JW, Creswell JD. Research design: Qualitative, quantitative, and mixed methods approaches. [Online]. Thousand Oaks: SAGE Publications; 2022 [cited 2025 Julio 13. Available from: https://spada.uns.ac.id/pluginfile.php/510378/mod_resource/content/1/creswell.pdf.
31. Brutsaert EF. MANUAL MSD. [Online].; 2025 [cited 2025 Julio 21. Available from: <https://www.msmanuals.com/es/hogar/trastornos-hormonales-y-metab%C3%B3licos/diabetes-mellitus-y-otros-trastornos-del-metabolismo-de-la-glucosa-sangu%C3%ADnea/diabetes-mellitus>.
32. American Diabetes Association Professional Practice Committee. Summary of Revisions: Standards of Care in Diabetes—2024. [Online].; 2024. Available from: <https://doi.org/10.2337/dc24-S001>.

33. Real Academia Española. Diccionario panhispánico del español jurídico, 2025. [Online].; 2025. Available from: <https://dpej.rae.es/lema/edad>.
34. Organización Mundial de la Salud. Gender and health (tema de edad, página general sobre género y salud). [Online].; 2025. Available from: <https://www.who.int/health-topics/gender>.
35. Instituto Nacional de Estadística e Informática. Encuesta Demográfica y de Salud Familiar (ENDES) 2022. [Online]. Lima: INEI; 2023 [cited 2025 Julio 20. Available from: <https://www.inec.es/DEFIne/concepto.htm?c=4484>.
36. Organización Mundial de la Salud. Gender and health (hoja informativa sobre género). [Online].: Organización Mundial de la Salud; 2025. Available from: <https://www.who.int/news-room/fact-sheets/detail/gender>.
37. Organización Panamericana de la Salud. OPS. [Online].; 2025. Available from: <https://www.paho.org/es/temas/prevencion-obesidad#:~:text=Un%20IMC%20igual%20o%20superior,adultos%20de%20todas%20las%20edades>.
38. World Health Organization Consultation. Obesity: Preventing and Managing the Global Epidemic; Report of a WHO Consultation. [Online].: World Health Organization. Available from: <https://iris.who.int/handle/10665/42330>.
39. DeMers D, Wachs D. National Library for Medicine. [Online].; 2025. Available from: [https://www.ncbi.nlm.nih.gov/books/NBK538226/#:~:text=La%20presi%C3%B3n%20arterial%20media%20\(PAM\)%20funciona%20para%20perfundir%20sangre%20a,eficazmente%20a%20todos%20los%20tejidos](https://www.ncbi.nlm.nih.gov/books/NBK538226/#:~:text=La%20presi%C3%B3n%20arterial%20media%20(PAM)%20funciona%20para%20perfundir%20sangre%20a,eficazmente%20a%20todos%20los%20tejidos).
40. Whelton K, Carey RM, Aronow WS, Casey Jr DE, Collins KJ, Himmelfarb D, et al. Guideline for the Prevention, Detection, Evaluation, and Management of High Blood Pressure in Adults: A Report of the ACC/AHA Task Force on Clinical Practice Guidelines. [Online].: Artículo de revista científica (guía clínica del Task Force ACC/AHA). Available from: <https://doi.org/10.1161/HYP.0000000000000065>.
41. MedlinePlus. MedlinePlus (servicio de la Biblioteca Nacional de Medicina de EE. UU.). [Online].; 2023. Available from: <https://medlineplus.gov/spanish/a1c.html>.

42. Sandeep K. Dhaliwal MbcIDEaMSVArbDCDMMDBC. MedlinePlus (enciclopedia médica, servicio de la Biblioteca Nacional de Medicina de EE.UU.). [Online].; 2025. Available from: https://medlineplus.gov/spanish/ency/esp_imagepages/19723.htm.
43. National Institute of Diabetes and Digestive and Kidney Diseases. Diabetes Tests & Diagnosis. NIDDK – Diabetes Diagnosis. .
44. Mayo Clinic. Mayo Clinic (versión en español). [Online].; 2024. Available from: <https://www.mayoclinic.org/es/tests-procedures/glucose-tolerance-test/about/pac-20394296>.
45. Qing Qiao. Definition and Diagnosis of Diabetes Mellitus and Intermediate Hyperglycaemia: Report of a WHO/IDF Consultation. [Online].: World Health Organization. Available from: <https://www.who.int/publications/i/item/9241594934>.
46. Galán Cotrina V. Aplicación de la Metodología Crisp-DM a un proyecto de minería de datos en el entorno universitario. [Online]. Madrid, España: Repositorio Institucional UC3M [cited 2025 Julio 13. Available from: <https://e-archivo.uc3m.es/rest/api/core/bitstreams/714c5452-962e-44cf-993f-ebb3088d4aa5/content>.
47. American Diabetes Association. Diabetes diagnosis & tests. [Online].; 2025. Available from: https://diabetes.org/about-diabetes/diagnosis?utm_source=chatgpt.com.

AGRADECIMIENTO

Queremos expresar nuestro más sincero agradecimiento a nuestros asesores y colaboradores que brindaron su respaldo técnico y académico a lo largo de esta investigación. De manera especial, damos las gracias al Dr. Nilthon Arce Fernández y al Dr. Guillermo Núñez Sánchez. Su valiosa orientación, la disposición para compartir sus conocimientos con nosotras y su motivación constante fueron pieza clave para adentrarnos a explorar nuevas fronteras en el diagnóstico de la Diabetes mediante Inteligencia Artificial.

Asimismo, extendemos un profundo reconocimiento a nuestra casa superior de estudios y el Hospital San Javier de Bellavista. Al facilitarnos el acceso a sus recursos tecnológicos y académicos, transformaron este proyecto en una realidad tangible. Un agradecimiento especial a nuestros familiares y amigos; ellos han sido un pilar fundamental y el motor detrás de nuestro compromiso académico.

Por último, no queremos dejar de reconocer el apoyo incondicional de todas las personas que colaboraron con nosotras. Sus aportes, observaciones y críticas constructivas, enriquecieron las distintas etapas de este proceso, ayudándonos a cumplir con el objetivo trazado.

Las investigadoras

DEDICATORIA

Con profunda gratitud, dedicamos este trabajo a Dios, quien nos proveyó de la salud, la motivación y las fuerzas necesarias para mantenernos firmes y lograr superar cada uno de los obstáculos académicos. Asimismo, a nuestros padres, cuyo amor, esfuerzo incansable y ejemplo constante nos guiaron paso a paso. A nuestros hermanos y amigos, gracias por ser ese refugio indispensable en los momentos más complejos y por sus palabras de aliento constante. Finalmente, a todas las personas que apostaron por nosotras y sumaron fuerzas para materializar esta meta, les dedicamos esta investigación. Lo hacemos con la certeza de que servirá como fuente de inspiración y como cimiento sólido para próximos estudios.

ANEXOS.

ANEXO 1: OPERACIONALIZACIÓN DE VARIABLES

Variable	Definición conceptual	Definición operacional	Tipo	Dimensión	Indicador	Ítem	Técnica de medición	Instrumento
Presencia de Diabetes Mellitus	Es un trastorno en el cual el organismo no produce suficiente cantidad de insulina o no responde normalmente a la misma, lo que provoca concentraciones elevadas de glucosa en sangre. ⁽³¹⁾	Confirmación del diagnóstico médico de diabetes tipo 1 o tipo 2 registrada en la historia clínica ⁽³²⁾ .	Dependiente / Cualitativa politómica	No aplica	Diagnóstico médico	Condición actual: <ul style="list-style-type: none"> • 0 = No tiene diabetes • 1 = Prediabetes • 2 = Diabetes 	Revisión de historia clínica	Ficha de recolección de datos
Edad	Tiempo vivido por una persona expresada en años ⁽³³⁾	Número total de años cumplidos por el paciente al momento del estudio, obtenido	Independiente / Cuantitativa continua	No aplica	Edad cronológica	Edad en años cumplidos	Revisión de historia clínica	Ficha de recolección de datos

		de la historia clínica ⁽³⁴⁾ .						
Sexo	Según la OMS hace referencia a las características biológicas y fisiológicas que definen a hombres y mujeres ⁽³⁵⁾ .	Clasificación binaria del paciente como masculino o femenino, registrada en la historia clínica ⁽³⁶⁾ .	Independiente / Cualitativa dicotómica	No aplica	Hombre / Mujer	Masculino (M) / Femenino (F)	Revisión de historia clínica	Ficha de recolección de datos
Índice de masa corporal (IMC)	Es un índice simple de peso para la talla que se utiliza comúnmente para clasificar el sobrepeso y la obesidad en adultos. Se calcula dividiendo el peso de una persona en kilogramos por el cuadrado de la	Resultado del cálculo del peso (kg) / estatura ² (m ²), medido con báscula y tallímetro. Clasificación según rangos de la OMS (normal, sobrepeso, obesidad) ⁽³⁸⁾ .	Independiente / Cuantitativa continua	Estado nutricional	Relación peso/talla ²	Peso: Kg / Talla: m (se calcula IMC)	Cálculo a partir de peso y talla	Ficha + instrumentos antropométricos

	altura en metros. (37)							
Presión arterial promedio	Se define como la presión arterial promedio a lo largo de un ciclo cardíaco, la sístole y la diástole. La PAM se ve influenciada por el gasto cardíaco y la resistencia vascular sistémica, cada uno de los cuales se ve afectado por diversas variables. (39)	Promedio aritmético de las mediciones sistólica y diastólica (PAM = PAD + 1/3 (PAS – PAD)), tomada de registros clínicos o con esfigmomanómetro (40).	Independiente / Cuantitativa continua	Presión sistólica y diastólica	Valores en mmHg	Presión Arterial: Sistólica / diastólica mmHg	Revisión de historia clínica	Ficha de recolección de datos
HbA1c (Hemoglobina glicosilada)	Es un examen de sangre. Mide el nivel promedio de glucosa o azúcar en la sangre durante los últimos tres meses. (41)	Porcentaje de hemoglobina ligada a glucosa en sangre, medido mediante prueba de laboratorio; indica control glucémico promedio de los	Independiente / Cuantitativa continua	Glucemia promedio 3 meses	Porcentaje (%)	Resultado de HbA1c %	Examen de laboratorio	Resultados de laboratorio clínico

		últimos 3 meses (32).						
Glucemia en ayunas	La prueba de glucosa en ayunas es la forma más rápida de medir la glucosa en la sangre para el diagnóstico de diabetes. En ayunas significa que la persona no ha comido ni bebido nada (excepto agua) durante 8 a 12 horas antes del examen. El diagnóstico de diabetes se hace en una persona si su nivel de glucosa en la sangre es de 126 mg/dl o superior en dos pruebas separadas. (42)	Concentración de glucosa (mg/dL) en sangre luego de 8-12h sin ingesta calórica, registrada mediante análisis de laboratorio. Diagnóstico si ≥ 126 mg/dL en dos ocasiones (43).	Independiente / Cuantitativa continua	Glicemia basal	Concentración de glucosa en sangre	Resultado de glucosa en ayunas: mg/dL	Examen de laboratorio	Resultados de laboratorio clínico

Prueba de tolerancia oral a la glucosa (PTOG)	La prueba de tolerancia a la glucosa mide la respuesta del cuerpo al azúcar o glucosa. Este análisis también recibe el nombre de prueba oral de tolerancia a la glucosa. ⁽⁴⁴⁾	Medición de la glucosa en sangre dos horas después de la ingesta de 75g de glucosa, usada para evaluar tolerancia alterada a la glucosa o diabetes ⁽⁴⁵⁾ .	Independiente / Cuantitativa continua	Metabolismo glucosa	Glucemia a las 2h tras carga de glucosa	Resultado de PTOG (2h): mg/dL	Examen de laboratorio	Resultados de laboratorio clínico

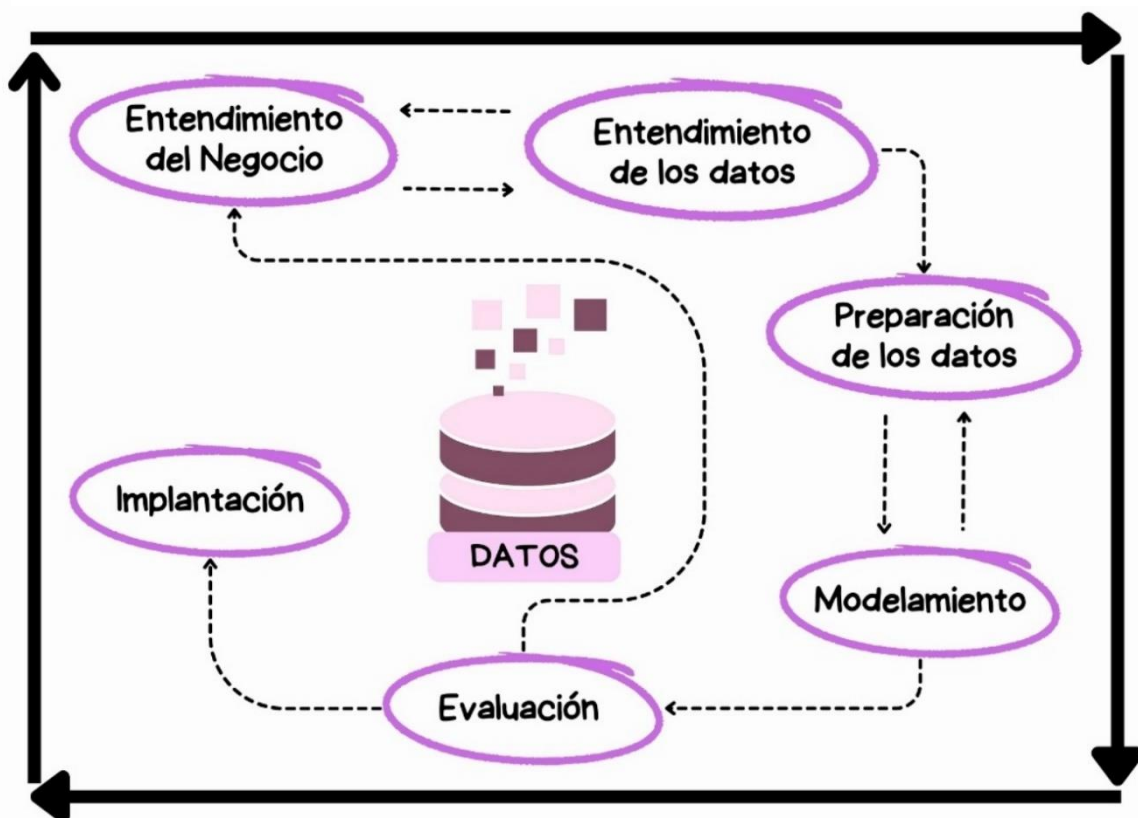
ANEXO 2: Metodología CRISP-DM

Fases para la construcción del Modelo de Red Neuronal Artificial

El modelo CRISP-DM (Cross Industry Standard Process for Data Mining), se constituye un estándar internacionalmente utilizado para orientar el proceso de minería de datos aplicada a la investigación. El enfoque utilizado por CRISP-DM se aplica de manera transversal en diversas industrias y disciplinas, proporciona una metodología estructurada que ayuda a los equipos de ciencia de datos a desarrollar proyectos de manera eficaz ⁽⁴⁶⁾.

Las fases en las que se divide CRISP-DM y las posibles secuencias a seguir entre ellas.

Figura 3. Secuencia del proceso CRISP-DM. Adaptado de Galán Cotrina V (UC3M) ⁽⁴⁶⁾



A. Entendimiento del Negocio: En la provincia de Jaén, la atención eficaz y oportuna de enfermedades como la diabetes mellitus representa un gran desafío, esto debido a la limitada disponibilidad de falta de recursos económicos. Al ver esta realidad de cerca, ha motivado la búsqueda de alternativas innovadoras y accesibles para el diagnóstico

temprano de dicha enfermedad, por ello, frente a esta situación decidimos desarrollar un modelo de red neuronal basado en inteligencia artificial (IA), el cual permite detectar de forma precisa y rápida esta enfermedad, utilizando los datos que hemos recopilado de manera local. Siendo nuestro objetivo el brindar una alternativa accesible que permita identificar casos de manera oportuna, convirtiéndose en un apoyo al diagnóstico mediante la identificación de casos tempranos, y así prevenir complicaciones graves, mejorando la salud y calidad de vida de quienes padecen esta patología en nuestra provincia ⁽⁴⁶⁾.

B. Entendimiento de los datos: Con el objetivo de aplicar algoritmos de inteligencia artificial en la predicción de la diabetes mellitus en la provincia de Jaén, se llevó a cabo inicialmente un proceso sistemático de recopilación y organización de datos clínicos a nivel individual. Para ello, se seleccionaron variables biomédicas relevantes entre ellas edad, género, índice de masa corporal (IMC), presión arterial promedio, glicemia basal, hemoglobina glicosilada y resultados de prueba de tolerancia oral a la glucosa (PTOG). Para organizar la información, cada registro recibió un identificador único por paciente, lo que garantizó una gestión estructurada y trazable de la base de datos desde las primeras etapas ⁽⁴⁶⁾.

Este abordaje no solo facilitó el análisis del contexto epidemiológico local, sino que a su vez permitió realizar una revisión detallada de la estructura, coherencia y completitud de los datos disponibles. Al momento de asegurar la calidad, aplicamos una depuración minuciosa (centrada en la identificación y corrección de los valores faltantes, inconsistencias y duplicaciones). Con esta limpieza buscábamos disponer de un conjunto robusto y a su vez confiable, ideal para el entrenamiento del modelo predictivo ⁽⁴⁶⁾.

C. Preparación de los datos: Tras finalizar la recolección de la información clínica, procedimos a organizar y depurar los datos para garantizar su calidad antes de introducirlos al modelo predictivo. Para ello, utilizamos como herramienta de apoyo Microsoft Excel® 2021, lo cual resultó clave para manejar los registros de manera clara, ordenada y eficiente; agilizando a su vez la detección de inconsistencias y la identificación de relaciones entre variables. Consideramos que esta fase resultó

esencial para garantizar que los datos enviados al modelo se encontraran completos y estructurados de manera sistemática, libre de duplicaciones⁽⁴⁶⁾.

Debemos tener en cuenta que lograr estas condiciones es indispensable para un aprendizaje efectivo del modelo, pues de la calidad de estos datos depende su capacidad para detectar patrones significativos y representativos ⁽⁴⁶⁾.

Una vez validada la información, se procedió a dividir en dos el conjunto de datos: el 80 % para el entrenamiento del modelo, con el propósito de que este orientado al reconocimiento de patrones y el 20 % restante para la validación, evaluando así el desempeño del modelo y comprobando su adecuación a los objetivos de la investigación ⁽⁴⁶⁾.

D. Modelamiento: En esta etapa, una vez que se obtuvo los datos organizados, esta base fue llevada a Python, un lenguaje de programación que nos permitió la creación del modelo de red neuronal, para ello utilizamos herramientas especializadas que nos permiten armar la estructura del modelo paso a paso, bibliotecas como TensorFlow y Scikit-learn. Aquí se define cuántas neuronas tendría en la capa de entrada, donde el modelo recibe toda la información inicial, cuántas serían necesarias en las capas ocultas, que son los niveles de procesamiento donde la base de datos es analizada y se refina, y finalmente cuántas neuronas incluiríamos en la capa de salida, que es donde el modelo nos da su respuesta al objetivo del estudio, la detección a la Diabetes Mellitus ⁽⁴⁶⁾.

Durante este proceso se realizó múltiples configuraciones: ajustamos la cantidad y modificamos la tasa de aprendizaje, que es la velocidad con la que el modelo va absorbiendo el conocimiento y realizando a su vez los ajustes a las funciones de activación, las cuales son determinantes para que las neuronas se enciendan o apaguen durante el momento del procesamiento de la información, entre otros detalles ⁽⁴⁶⁾.

E. Evaluación del modelo: Garantizar que el modelo de red neuronal desarrollado no solo tuviera un buen desempeño con los datos de entrenamiento, sino que también pudiera generalizar adecuadamente ante nuevos casos en contextos reales, se implementaron técnicas de regularización y validación cruzada. Estas metodologías

permitieron someter al modelo a múltiples particiones de los datos, evaluando su robustez y capacidad de adaptación frente a pacientes con características distintas a las del conjunto de entrenamiento ⁽⁴⁶⁾.

Con el propósito de evaluar el desempeño del modelo se recurrió a métricas reconocidas en problemas de clasificación, entre ellas la precisión (accuracy), la sensibilidad (recall), la especificidad y la curva ROC (Receiver Operating Characteristic), que ofrece visualizar gráficamente el comportamiento del modelo en distintos puntos de corte. Además, el análisis se reforzó incorporando indicadores de error como el error cuadrático medio (MSE) y el coeficiente de determinación (R2) los cuales fueron útiles para comprobar la fiabilidad de las predicciones y detectar áreas en las que el modelo podría ajustarse mejor ⁽⁴⁶⁾.

- F. **Implantación del Modelo:** Luego de haber realizado la evaluación, se propone la integración del modelo en plataformas digitales de salud, lo cual favorecerá la automatización en el diagnóstico y la detección temprana de la Diabetes Mellitus. Para llevar a cabo esta etapa, la cooperación de organizaciones de salud para asegurar una adecuada aplicación y formación del personal en el manejo de esta nueva herramienta que será de gran utilidad ⁽⁴⁶⁾.

Uno de los principales desafíos al momento de trabajar con redes neuronales artificiales es la limitada interpretabilidad de sus resultados, incluso cuando se muestran altos niveles de precisión, por ello para resolver esta dificultad y facilitar su comprensión a los profesionales de la salud, se optó por técnicas de interpretación como SHAP (Shapley Additive exPlanations); dichos métodos nos permiten visualizar la contribución de cada variables en la predicción final, lo cual facilita su comprensión y aumenta la confianza en el modelo ⁽⁴⁶⁾.

Librerías para procesar los datos

A. Instalar la librería pandas-datareader

Se instaló la librería “*pandas-datareader*” mediante el comando *pip install pandas-datareader*. El resultado de la acción fue la habilitación de funciones para la obtención y organización eficiente de datos clínicos y epidemiológicos, integrados al entorno de

pandas. Este procedimiento nos permitió perfeccionar el preprocesamiento del dataset, garantizando que la información utilizada en el entrenamiento y validación del modelo predictivo de la diabetes mellitus fuese consistente, reproducible y confiable.

B. Instalar la librería TensorFlow

A nivel de software, el primer paso fue la instalación de la librería “*TensorFlow*” (empleando el comando `pip install tensorflow`). El resultado directo de esta acción fue contar con un entorno especializado para el diseño, ideal para diseñar, entrenar y validar nuestras redes neuronales artificiales. Esto optimizó por completo el trabajo con los algoritmos de aprendizaje profundo. Viéndolo en retrospectiva, esta herramienta resultó fundamental para la construcción del modelo predictivo de la diabetes mellitus, al permitir procesar sin problemas los grandes volúmenes de datos clínicos, ajustar parámetros clave y evaluar el desempeño general del sistema. Garantizando solidez y confiabilidad en los resultados obtenidos.

C. Importar las librerías NumPy, Pandas y Matplotlib

En cuanto a la gestión de la información, nos apoyamos en librerías como *Matplotlib* y *NumPy*, *Pandas*, las cuales nos brindaron las herramientas necesarias para abordar el procesamiento numérico, la estructuración de las bases clínicas y la posterior visualización de los datos relevantes. Contar con este entorno de trabajo, resultó verdaderamente determinante al momento de realizar el análisis exploratorio y preparar el dataset adecuadamente. Fue un paso determinante para poder construir y aplicar el modelo de RNA en la predicción de la diabetes mellitus en la provincia de Jaén.

D. Importar las librerías IPython, IPython.display y Matplotlib (como mpl)

Contribuyó de manera específica al logro del objetivo de construir y evaluar un modelo de red neuronal artificial para predecir diabetes en Jaén. La importación de librerías proporcionó las herramientas necesarias para el manejo de datos, la visualización de resultados y la construcción de modelos. La normalización con `MinMaxScaler` permitió estandarizar las variables clínicas y mejorar la eficiencia del entrenamiento. La incorporación de arquitecturas neuronales mediante TensorFlow/Keras (Dense, Dropout, GRU, LSTM, Conv1D) posibilitó el reconocimiento de relaciones complejas entre los indicadores de salud analizados. Finalmente, la configuración de parámetros gráficos generó visualizaciones claras que facilitaron la interpretación de tendencias y la validación del desempeño del modelo predictivo propuesto.

Importando la base de datos

El bloque de código utilizado para importar y explorar el archivo DATA_FINAL.xlsx accediendo a la hoja Base_de_Datos y verificando sus primeros registros, aseguró la carga correcta y validación inicial de los datos clínicos. Este procedimiento, integrado con las demás librerías de procesamiento y modelado, constituyó el primer paso esencial para garantizar la confiabilidad, consistencia y organización de la información empleada en el preprocesamiento y posterior entrenamiento de RNA, lo que fortaleció la solidez de los resultados obtenidos en la predicción de la diabetes mellitus en Jaén.

Tabla 2. Registro de datos observados del Hospital – San Javier de Bellavista, ciudad de Jaén

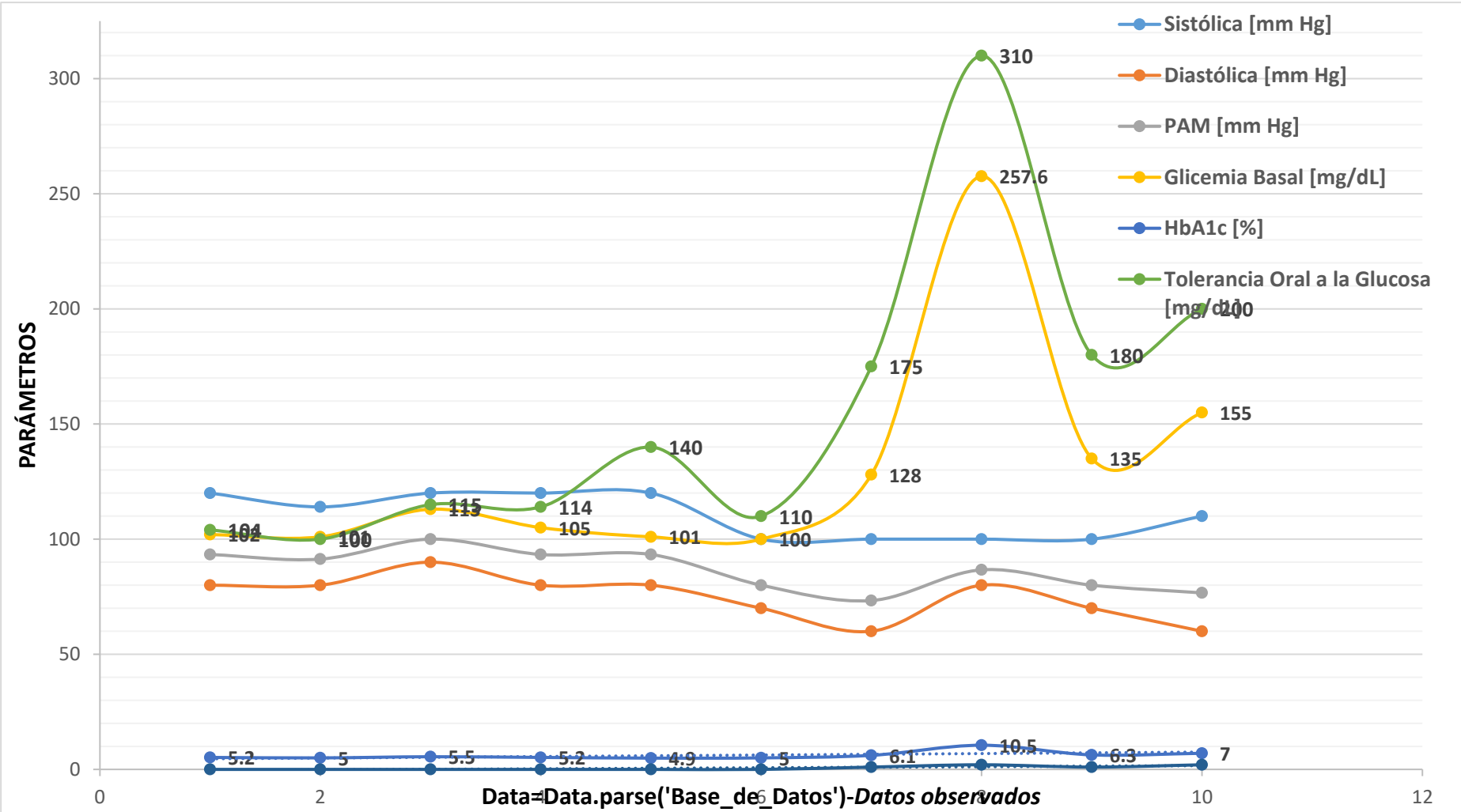
Edad	Sexo	IMC [kg/m ²]	Sistólica [mm Hg]	Diastólica [mm Hg]	PAM [mm Hg]	Glicemia Basal [mg/dL]	HbA1c [%]	Tolerancia Oral a la Glucosa [mg/dL]	Diagnóstico
28	F	46.4	120	80	93.33	102	5.2	104	0
50	F	30.2	114	80	91.33	101	5	100	0
52	F	33.9	120	90	100	113	5.5	115	0
58	M	28.8	120	80	93.333	105	5.2	114	0
61	F	21.1	120	80	93.33	101	4.9	140	0
60	F	27.5	100	70	80	100	5	110	0
49	M	30.7	100	60	73.33	128	6.1	175	1
54	F	27.5	100	80	86.66	257.6	10.5	310	2
51	F	25.3	100	70	80	135	6.3	180	1
61	F	24.9	110	60	76.66	155	7	200	2

Descripción e interpretación: Al revisar los registros clínicos presentados en la Tabla 1, se observa que los pacientes clasificados con diagnóstico 0 (es decir, sin diabetes) tienden a mostrar valores dentro de los rangos considerados normales en las principales variables metabólicas, como es la glucosa en ayunas, hemoglobina glicosilada (HbA1c) y la prueba de tolerancia oral a la glucosa. En este grupo, los niveles de glucosa en ayunas se sitúan entre 100 y 113 mg/dL, mientras que los niveles de HbA1c se encuentran entre 4.9 % y 5.5 %, lo que nos sugiere un buen equilibrio metabólico. Además, los resultados de la prueba de tolerancia oral se mantienen por debajo de los 140 mg/dL, lo que refuerza la idea de un metabolismo de la glucosa dentro de parámetros saludables.

Por el contrario, los pacientes con diagnóstico 1 o 2 presentan indicadores clínicos alterados. Un caso representativo es el de un paciente con diagnóstico 2, quien muestra glicemia basal de 257.6 mg/dL, la HbA1c de 10.5 % y la tolerancia oral a la glucosa de 310 mg/dL valores considerablemente superiores a lo esperado y consistentes con un descontrol metabólico severo.

De manera similar los demás pacientes con diagnóstico positivo (1 o 2) muestran niveles de glucosa en ayunas iguales o superiores a 128 mg/dL y valores de HbA1c que oscilan entre 6.5 % y 7 %, lo que nos indica una alteración persistente en el manejo de glucosa. Cabe señalar que, aunque algunos individuos sin diagnóstico presentan índices de masa corporal elevados incluso superiores a 3^o kg/ m², no manifiestan alteraciones en los parámetros metabólicos evaluados.

Figura 4. Estadígrafo del registro de datos observados del Hospital San Javier de Bellavista, ciudad de Jaén



Análisis de los valores de las variables

Tabla 3. Estadísticos descriptivos de los datos en Edad

Parámetros		Estadístico	Error estándar	
Edad	Media	52,7157	,54578	
	95% de intervalo de confianza para la media	Límite inferior	51,6428	
		Límite superior	53,7885	
	Media recortada al 5%	53,2095		
	Mediana	55,0000		
	Varianza	123,619		
	Desviación estándar	11,11843		
	Mínimo	18,00		
	Máximo	70,00		
	Rango	52,00		
	Rango intercuartil	16,00		
	Asimetría	-,643	,120	
	Curtosis	-,200	,239	

Descripción e interpretación: En relación con la variable edad de los participantes, se observó una media de 52,72 años, con un error estándar de 0,55, lo que indica precisión en la estimación de la media poblacional. El intervalo de confianza al 95% para la media se encuentra entre 51,64 y 53,79 años, lo que respalda esta estimación. La mediana fue de 55 años, superior a la media, mientras que la media recortada al 5 % fue de 53,21, lo cual sugiere la presencia de valores bajos atípicos. La desviación estándar fue de 11,12, con una varianza de 123,62, reflejando una dispersión moderada en los datos.

El rango total de edades fue de 52 años, con un mínimo de 18 y un máximo de 70 años, mientras que el rango intercuartílico fue de 16 años, mostrando concentración en los valores centrales. En cuanto a la forma de distribución, se identificó una asimetría negativa de -0,643 con un error estándar de 0,120, lo que indica una ligera inclinación hacia la izquierda (mayoría de edades por encima de la media), mientras que la curtosis fue de -0,200 (error estándar 0,239), lo que sugiere una distribución levemente más plana que la normal. Estos resultados permiten caracterizar a la población como predominantemente adulta, con una distribución de edades relativamente homogénea y sin una dispersión extrema.

Tabla 4. Estadísticos descriptivos de los datos del IMC

Parámetros			Estadístico	Error estándar
IMC [kg/m ²]	Media		28,3880	,20578
	95% de intervalo de confianza para la media	Límite inferior	27,9835	
		Límite superior	28,7925	
	Media recortada al 5%		28,2083	
	Mediana		27,7000	
	Varianza		17,573	
	Desviación estándar		4,19201	
	Mínimo		16,90	
	Máximo		46,40	
	Rango		29,50	
	Rango intercuartil		4,80	
	Asimetría		,766	,120
	Curtosis		1,152	,239

Descripción e interpretación: Respecto al Índice de Masa Corporal (IMC) de los participantes, se obtuvo una media de 28,39 kg/m², con un error estándar de 0,21, lo que indica una estimación precisa de la media poblacional. El intervalo de confianza al 95 % para la media se sitúa entre 27,98 y 28,79 kg/m², reafirmando esta precisión. La mediana fue de 27,70 kg/m², ligeramente inferior a la media, y la media recortada al 5 % fue de 28,21 kg/m², lo que sugiere la presencia de algunos valores extremos altos. La desviación estándar fue de 4,19 y la varianza de 17,57, evidenciando una variabilidad moderada en los datos.

El rango total de los valores fue de 29,50 kg/m², con un mínimo de 16,90 y un máximo de 46,40 kg/m², mientras que el rango intercuartílico fue de 4,80, lo que muestra una cierta concentración de valores en el centro de la distribución. La asimetría positiva de 0,766 (error estándar de 0,120) indica que la distribución está sesgada hacia la derecha, es decir, existen participantes con valores de IMC elevados.

Asimismo, la curtosis de 1,152 (error estándar de 0,239) revela una distribución más apuntada que la normal (leptocúrtica), lo que implica mayor concentración de valores alrededor de la media y presencia de valores extremos. En conjunto, estos resultados sugieren una tendencia general al sobrepeso en la población evaluada, con algunos casos que podrían encontrarse en rangos de obesidad, según los criterios de clasificación.

Tabla 5. Estadísticos descriptivos de los datos del PAM

Parámetros		Estadístico	Error estándar	
PAM [mm Hg]	Media	91,3297	,01059	
	95% de intervalo de confianza para la media	Límite inferior	91,3088	
		Límite superior	91,3505	
	Media recortada al 5%	91,3065		
	Mediana	91,2860		
	Varianza	,047		
	Desviación estándar	,21568		
	Mínimo	91,10		
	Máximo	93,30		
	Rango	2,20		
	Rango intercuartil	,19		
	Asimetría	2,936	,120	
	Curtosis	17,371	,239	

Descripción e interpretación: En cuanto a la Presión Arterial Media (PAM), se obtuvo una media de 91,33 mm Hg, con un error estándar de 0,011, lo que refleja una alta precisión en la estimación. El intervalo de confianza al 95 % para la media se encuentra entre 91,31 y 91,35 mm Hg, indicando estabilidad en los valores promedio registrados. La mediana fue de 91,29 mm Hg, y la media recortada al 5 % fue similar (91,31 mm Hg), lo que sugiere una distribución central consistente, aunque con presencia de valores extremos. La desviación estándar fue de solo 0,22 mm Hg, con una varianza de 0,047, evidenciando una variabilidad extremadamente baja entre los datos. El rango total fue de 2,20 mm Hg, con un mínimo de 91,10 y un máximo de 93,30 mm Hg, mientras que el rango intercuartílico fue de apenas 0,19 mm Hg, lo que indica una distribución muy compacta.

No obstante, la asimetría positiva de 2,936 (error estándar 0,120) revela una distribución fuertemente sesgada hacia la derecha, es decir, con presencia de algunos valores más altos. Asimismo, la curtosis de 17,371 (error estándar 0,239) indica una distribución altamente leptocúrtica, muy apuntada, con valores concentrados en torno a la media y presencia de valores extremos. En conjunto, aunque los valores de PAM se muestran muy consistentes en la mayoría de los casos, la elevada asimetría y curtosis evidencian una distribución anormal, influenciada por unos pocos valores atípicos en el extremo superior.

Tabla 6. Estadísticos descriptivos de los datos de la Glicemia Basal

Parámetros			Estadístico	Error estándar
Glicemia Basal [mg/dL]	Media		131,7588	2,96115
	95% de intervalo de confianza para la media	Límite inferior	125,9380	
		Límite superior	137,5795	
	Media recortada al 5%		123,1205	
	Mediana		109,0000	
	Varianza		3638,900	
	Desviación estándar		60,32329	
	Mínimo		78,00	
	Máximo		470,30	
	Rango		392,30	
	Rango intercuartil		34,83	
	Asimetría		2,523	,120
	Curtosis		6,753	,239

Descripción e interpretación: En relación con la glicemia basal de los participantes, se obtuvo una media de 131,76 mg/dL, con un error estándar de 2,96, lo que refleja una variabilidad importante en la estimación. El intervalo de confianza al 95 % para la media se ubica entre 125,94 y 137,58 mg/dL, lo cual indica cierta dispersión en los valores registrados. La mediana fue de 109,00 mg/dL, notablemente menor que la media, y la media recortada al 5 % fue de 123,12 mg/dL, lo que evidencia la presencia de valores extremos altos que están influyendo en el promedio. La desviación estándar fue de 60,32 mg/dL, con una varianza de 3.638,90, reflejando una alta dispersión de los datos.

El valor mínimo registrado fue de 78,00 mg/dL y el máximo de 470,30 mg/dL, generando un rango total de 392,30 mg/dL, mientras que el rango intercuartílico fue de 34,83 mg/dL, indicando dispersión moderada entre el 25 % y 75 % central de los casos. La asimetría positiva de 2,523 (error estándar de 0,120) indica una fuerte inclinación hacia la derecha, es decir, predominan valores más bajos, pero existen varios valores atípicos altos. La curtosis de 6,753 (error estándar de 0,239) señala una distribución leptocúrtica, con una mayor concentración de valores cercanos a la media y presencia de valores extremos.

En conjunto, estos resultados muestran que, si bien la media supera el umbral de Normo glicemia (100 mg/dL), la distribución se ve fuertemente influida por valores

atípicos de hiperglicemia, lo cual sugiere la posible presencia de casos con alteraciones metabólicas o diabetes no controlada dentro del grupo estudiado.

Tabla 7. Estadísticos descriptivos de los datos de la Hemoglobina Glicosilada

Parámetros		Estadístico	Error estándar
HbA1c [%]	Media	6,4792	,09097
	95% de intervalo de confianza para la media	Límite inferior	6,3004
		Límite superior	6,6580
	Media recortada al 5%		6,2417
	Mediana		5,8000
	Varianza		3,434
	Desviación estándar		1,85316
	Mínimo		4,00
	Máximo		15,90
	Rango		11,90
	Rango intercuartil		1,20
	Asimetría		2,247
			,120
	Curtosis		5,095
			,239

Descripción e interpretación: En cuanto a los valores de hemoglobina glicosilada (HbA1c) de los participantes, se obtuvo una media de 6,48 %, con un error estándar de 0,091, lo que indica una estimación precisa. El intervalo de confianza al 95 % para la media se encuentra entre 6,30 % y 6,66 %, sugiriendo que el promedio real de la población estudiada se sitúa en un rango levemente superior al umbral de normalidad. La mediana fue de 5,80 %, menor que la media, y la media recortada al 5 % fue de 6,24 %, lo cual evidencia la influencia de valores extremos altos sobre la media general.

La desviación estándar fue de 1,85 %, con una varianza de 3,43, lo que muestra una variabilidad considerable en los valores de HbA1c. El valor mínimo registrado fue de 4,00 %, mientras que el máximo alcanzó el 15,90 %, generando un rango de 11,90 %; por su parte, el rango intercuartílico fue de 1,20 %, lo cual indica que los valores centrales están relativamente concentrados. Sin embargo, la asimetría positiva de 2,247 (error estándar 0,120) indica una distribución altamente sesgada hacia la derecha, es decir, con varios casos de HbA1c elevados.

Además, la curtosis de 5,095 (error estándar 0,239) muestra una distribución leptocúrtica, con alta concentración de valores en torno a la media y presencia significativa de valores extremos. En conjunto, estos resultados reflejan que, si bien una

parte importante de la población se encuentra dentro de rangos normales o prediabéticos, existe una proporción significativa con HbA1c en niveles compatibles con diabetes mal controlada, lo que podría alertar sobre una situación de riesgo metabólico en el grupo evaluado.

Tabla 8. Estadísticos descriptivos de los datos de la Tolerancia Oral a la Glucosa

Parámetros		Estadístico	Error estándar	
Tolerancia Oral a la Glucosa [mg/dL]	Media	169,4217	4,96430	
	95% de intervalo de confianza para la media	Límite inferior	159,6633	
		Límite superior	179,1801	
	Media recortada al 5%	155,9665		
	Mediana	138,0000		
	Varianza	10227,394		
	Desviación estándar	101,13058		
	Mínimo	61,00		
	Máximo	1197,00		
	Rango	1136,00		
	Rango intercuartil	66,00		
	Asimetría	4,826	,120	
	Curtosis	37,082	,239	

Descripción e interpretación: En lo referente a la Tolerancia Oral a la Glucosa (TOG), se obtuvo una media de 169,42 mg/dL, con un error estándar de 4,96, lo que sugiere una estimación razonablemente precisa. El intervalo de confianza al 95 % para la media se sitúa entre 159,66 y 179,18 mg/dL, lo que evidencia una dispersión considerable entre los valores. La mediana fue de 138,00 mg/dL, y la media recortada al 5 % fue de 155,97 mg/dL, indicando que la media general está influenciada por valores atípicos elevados. Esta tendencia se refuerza con la desviación estándar de 101,13 mg/dL y una varianza muy elevada de 10.227,39, lo que revela una altísima dispersión en los datos.

El valor mínimo registrado fue de 61,00 mg/dL y el máximo alcanzó los 1.197,00 mg/dL, resultando en un rango extremo de 1.136,00 mg/dL, mientras que el rango intercuartílico fue de 66,00 mg/dL, mostrando cierta variabilidad en los valores centrales. La asimetría positiva de 4,826 (error estándar 0,120) indica una distribución fuertemente sesgada hacia la derecha, es decir, existen múltiples casos con niveles de glucosa posprandial muy elevados.

Asimismo, la curtosis de 37,082 (error estándar 0,239) señala una distribución extremadamente leptocúrtica, con fuerte concentración de valores en torno a la media y presencia de valores extremos. En conjunto, los resultados reflejan una alta heterogeneidad metabólica en la muestra analizada, con presencia significativa de casos

que probablemente corresponden a diabetes mal controlada o no diagnosticada, dado que los valores normales de TOG deben estar por debajo de los 140 mg/dL según los criterios clínicos.

Estudio de la correlación de las variables

Se realizó el estudio de correlación de variables mediante el coeficiente correlación Pearson, donde se identificó las variables que más influyen en el diagnóstico de la enfermedad Diabetes Mellitus, y cuyos resultados se presentan en la Tabla 9.

Tabla 9. Estudio de correlación de variables mediante Pearson

	Edad	Sexo	IMC	Sistólica	Diastólica	PAM	Glicemia_B asal	HbA1c	Tolerancia_Or al_Glucosa	Diagnóstico
Edad	1	0.061242	-0.10267	0.205384	0.19584	0.213696	0.100851	0.137866	0.113722	0.154951
Sexo	0.061242	1	0.015053	0.005265	0.017045	0.011753	-0.034154	0.005812	-0.030118	0.027095
IMC	-0.10267	0.015053	1	0.000038	0.024032	0.012562	-0.077794	-0.05752	-0.097344	-0.03633
Sistólica	0.205384	0.005265	0.000038	1	0.764145	0.941642	0.094174	0.196004	0.122619	0.10789
Diastólica	0.19584	0.017045	0.024032	0.764145	1	0.936684	0.117498	0.190668	0.130245	0.119157
PAM	0.213696	0.011753	0.012562	0.941642	0.936684	1	0.112441	0.205905	0.134535	0.120752
Glicemia_Bas al	0.100851	-0.034154	-0.077794	0.094174	0.117498	0.112441	1	0.812156	0.864178	0.585994
HbA1c	0.137866	0.005812	-0.05752	0.196004	0.190668	0.205905	0.812156	1	0.832185	0.717574
Tolerancia_Or al_Glucosa	0.113722	-0.030118	-0.097344	0.122619	0.130245	0.134535	0.864178	0.832185	1	0.642614
Diagnóstico	0.154951	0.027095	-0.03633	0.10789	0.119157	0.120752	0.585994	0.717574	0.642614	1

Descripción e interpretación: Para consolidar nuestra propuesta del modelo de red neuronal artificial, el cual está enfocado a la predicción de la Diabetes Mellitus en la ciudad de Jaén durante el año 2023, se optó por ejecutar un análisis de correlación entre las variables clínicas observadas y los resultados emitidos por el modelo. El coeficiente de correlación permitió identificar relaciones significativas entre diversas variables fisiológicas y metabólicas con el diagnóstico final. En la revisión de los datos, destaca la fuerte correlación positiva entre la glicemia basal y la hemoglobina glicosilada (HbA1c) ($r = 0.812$), a la par de la relación entre la glicemia basal y la prueba de tolerancia oral a la glucosa ($r = 0.864$). En la práctica clínica, esto evidencia la consistencia de estos tres indicadores al momento de caracterizar el estado glucémico del paciente.

Siguiendo esta misma línea, la hemoglobina glicosilada también mostró una correlación considerable con el diagnóstico final ($r = 0.717$), al igual que la tolerancia oral a la glucosa ($r = 0.642$), confirmando su valor predictivo dentro del modelo propuesto. Por otro lado, al evaluar los parámetros hemodinámicos – como la presión sistólica, diastólica y la presión arterial media (PAM) –, notamos que presentaron una relación moderada con las variables glucémicas y con el diagnóstico, Este hallazgo nos sugiere una posible asociación entre la presión arterial y la presencia de alteraciones metabólicas.

Aunque el índice de masa corporal (IMC) presentó correlaciones negativas débiles con la mayoría de las variables, su participación dentro del modelo no puede descartarse, ya que podría influir de forma no lineal en la predicción. Por otro lado, variables como el sexo y la edad mostraron correlaciones bajas pero positivas con el diagnóstico, lo cual indica que, si bien no son determinantes por sí solas, podrían actuar como factores complementarios en la estructura del modelo neuronal. En conjunto, los resultados reflejan que las variables seleccionadas no solo tienen fundamentos clínicos, sino que también presentan interrelaciones que justifican su inclusión en el diseño del modelo predictivo basado en redes neuronales artificiales.

- **Correlación entre Edad y Diagnóstico**

El coeficiente de correlación de Pearson es $r = 0.154951$; correlación positiva débil. Aunque la relación es baja, indica que a mayor edad hay una ligera tendencia a

diagnosticar diabetes. La relación no es lo suficientemente fuerte como para considerarse significativa en este contexto.

Esta variable es importante ya que es un factor determinante en el riesgo de padecer diabetes mellitus tipo 2. Ya que pues con el paso del tiempo se reduce la sensibilidad a la insulina. Y esto, a su vez, aumentar la probabilidad de alteraciones metabólicas. Al incluir, esta variable en el modelo buscamos identificar cuáles son los patrones o tendencias relacionadas al envejecimiento y ajustar también las predicciones de este mismo de acuerdo con los diferentes grupos etario, la consideramos importante y útil en contextos de prevención y detección temprana.

- **Correlación entre Sexo y Diagnóstico**

El coeficiente de correlación de Pearson es $r = 0.027095$, correlación prácticamente nula. El sexo no presenta relación relevante con el diagnóstico en esta muestra.

Incluir esta variable en nuestro modelo nos permite considerar las diferencias biológicas y hormonales que existen entre hombres y mujeres, las cuales estas diferencias pueden influir. En el metabolismo de la glucosa es decir la distribución de la grasa corporal, y. La respuesta a la insulina. Además, diversos estudios han demostrado que la prevalencia a la diabetes puede variar según el sexo. Por lo que esta variable puede ayudar a la precisión y, a su vez, a la personalización del modelo de desarrollar.

- **Correlación entre IMC y Diagnóstico**

El coeficiente de correlación de Pearson es $r = -0.03633$; correlación negativa débil. El IMC, en este conjunto de datos, no muestra una relación clara con el diagnóstico.

Esta variable es considera ya que el IMC, al ser considerada una medida para evaluar el estado nutricional de una persona, se encuentra estrechamente relacionado con el riesgo de resistencia a la insulina y a la aparición de la diabetes. La excesiva acumulación de grasa abdominal particularmente en la zona del abdomen es un indicador como factor de riesgo. Por eso, el incorporar esta variable, permite al modelo tener una visión más clara del componente físico y su impacto en el metabolismo de la glucemia.

- **Correlación entre Presión Sistólica y Diagnóstico**

El coeficiente de correlación de Pearson es $r = 0.10789$; correlación positiva débil. Su relación con el diagnóstico es limitada.

La presión sistólica representa la fuerza con la que la sangre empuja contra las paredes de las arterias cuando el corazón se contrae. Valores elevados de Presión sistólica

han sido relacionados con un mayor riesgo de contraer y padecer enfermedades cardiovasculares y alteraciones metabólicas. Al incluir esta variable dentro del modelo de predicción de diabetes se obtiene una mejor comprensión de cómo la función cardiovascular se relaciona con el metabolismo de la glucosa. Su análisis, que me variable, permite detectar si la hipertensión sistólica sería un factor concurrente en los pacientes con niveles alterados de glicemia.

- **Correlación entre Presión Diastólica y Diagnóstico**

El coeficiente de correlación de Pearson es $r = 0.119157$, correlación positiva débil. No se evidencia una relación fuerte con el diagnóstico.

Considerando que la presión de Diastólica mide la presión en las arterias cuando el corazón está en reposo entre los latidos, aunque a veces esta suele recibir menos atención que la presión sistólica sus valores persistentemente elevados o ya sea bajos pueden reflejar problemas circulatorios y ciertas alteraciones en el equilibrio general del sistema cardiovascular. Al considerar esta variable nos permite identificar si hay asociación entre la presión basal del sistema circulatorio y el riesgo de desarrollar diabetes.

- **Correlación entre Presión Arterial Media (PAM) y Diagnóstico**

El coeficiente de correlación de Pearson es $r = 0.120752$; correlación positiva débil. No se demuestra una asociación relevante en este caso.

Una presión elevada suele estar relacionado con disfunciones metabólicas como la resistencia a la insulina y el síndrome metabólico. Al incluir esta variable en el modelo, nos permitiría encontrar la relación entre la regulación cardiovascular y el riesgo de padecer diabetes, permitiéndonos así encontrar factores combinados, que suele presentarse de forma simultánea en los pacientes.

- **Correlación entre Glicemia Basal y Diagnóstico**

El coeficiente de correlación de Pearson es $r = 0.585994$, que implica una correlación positiva moderada. Se considera clínicamente significativa.

La glicemia basal es una variable fundamental en el estudio, porque constituye una de los principales criterios diagnósticos de la diabetes según organismos internacionales. Debemos considerar que la glicemia basal elevada está directamente asociada a alteraciones metabólicas (como la intolerancia a la glucosa y la resistencia a la insulina), las cuales representan etapas iniciales en el desarrollo de la diabetes mellitus.

Al incluir esta variable en el modelo, reconocemos su rol diagnóstico y logramos establecer la relación entre el control glicémico y el riesgo de progresión hacia la enfermedad. En la práctica, esto nos permite identificar factores combinados y patrones clínicos que suelen presentarse simultáneamente en los pacientes, fortaleciendo la capacidad predictiva del análisis.

- **Correlación entre Hemoglobina Glicosilada (HbA1c) y Diagnóstico**

El coeficiente de correlación de Pearson es $r = 0.717574$, correlación positiva alta. refuerza su papel clave en el diagnóstico.

Dado que la hemoglobina glicosilada es un indicador clave del control glucémico a largo plazo (ya que refleja los niveles promedio de glucosa en la sangre durante aproximadamente los últimos 3 meses), resulta indispensable integrarla como variable en el modelo. Hacerlo nos permite evaluar la exposición prolongada del organismo del paciente a niveles elevados de glucosa, la cual fundamental para predecir riesgos metabólicos. A la par, resulta ser muy útil para detectar patrones de hiperglucemia sostenida.

- **Correlación entre Tolerancia Oral a la Glucosa y Diagnóstico**

El coeficiente de correlación de Pearson es $r = 0.642614$, correlación positiva considerable. Confirma su utilidad diagnóstica.

Considerando que la hemoglobina glicosilada es un indicador clave del control glucémico a largo plazo, ya que nos refleja los niveles promedio de glucosa en la sangre durante aproximadamente los últimos 3 meses. Es relevante considerarla, como variable en el modelo, ya que permite evaluar la exposición prolongada del organismo del paciente a niveles elevados de glucosa, la cual es fundamental para predecir riesgos metabólicos. Además, resulta ser útil para detectar patrones de hiperglucemia sostenida.

Seleccionando los datos de entrada y salida

En la propuesta de un modelo de red neuronal artificial para predecir la diabetes mellitus en la ciudad de Jaén–2023, se construyó una base de datos que integra variables clínicas de alta relevancia: Edad, Sexo, IMC, presión arterial sistólica y diastólica, presión arterial media (PAM), glicemia basal, HbA1c y tolerancia oral a la glucosa. Estas variables fueron procesadas como datos de entrada (Data_X), mientras que el diagnóstico médico se estableció como variable de salida (Data_Y). El aporte central de este resultado radica en que la red neuronal entrenada demuestra que, a partir de parámetros clínicos

básicos y rutinarios en los establecimientos de salud de Jaén, es posible generar predicciones confiables sobre el riesgo de diabetes mellitus.

Escalando los datos

En la investigación “Propuesta de un modelo de red neuronal artificial para predecir diabetes mellitus en la ciudad de Jaén–2023”, la aplicación del escalado de datos con MinMaxScaler permitió transformar las variables clínicas de entrada (Edad, Sexo, IMC, presión arterial, glicemia basal, HbA1c y tolerancia oral a la glucosa) y la variable de salida (Diagnóstico: 0 = No hay diabetes, 1 = Prediabetes, 2 = Diabetes). Esta normalización garantizó que el modelo procesara la información sin sesgos derivados de diferencias de escala, logrando un entrenamiento más estable y una mejor capacidad de generalización.

El aporte más relevante de este resultado consiste en que el modelo neuronal, al trabajar con datos normalizados, incrementó su precisión en la predicción de los tres estados de salud. Esto representa un avance significativo para el contexto de Jaén, donde la variabilidad clínica entre pacientes es alta, aportando una herramienta con potencial para apoyar el diagnóstico temprano y la prevención de la diabetes mellitus en la región.

Conjunto de entrenamiento y prueba

Se realizó la división de la base de datos en conjuntos de entrenamiento y prueba mediante la función `train_test_split`, destinando el 80% de los registros al entrenamiento del modelo y el 20% restante a la validación. Esta distribución permitió contar con un conjunto amplio de datos para el aprendizaje de la red neuronal y, al mismo tiempo, disponer de un grupo independiente para evaluar su desempeño en la predicción de los diagnósticos (0 = No hay diabetes, 1 = Prediabetes, 2 = Diabetes).

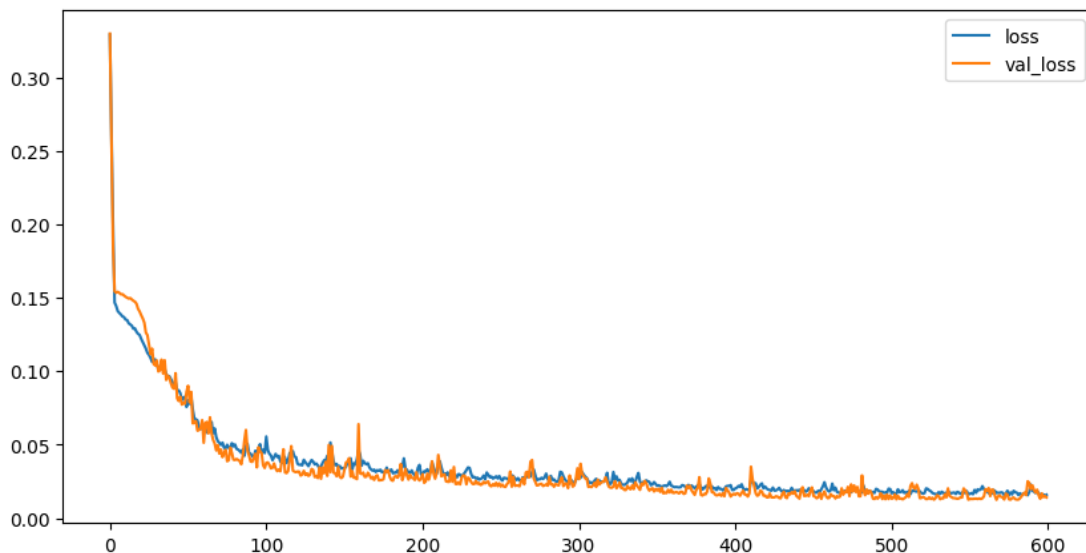
En el contexto de Jaén, esto resulta fundamental para garantizar que la herramienta propuesta tenga validez práctica en la detección temprana y clasificación del riesgo de diabetes mellitus, fortaleciendo la relevancia científica y aplicada de la investigación.

Construcción de un modelo de red neuronal (LSTM) para predecir Diabetes Mellitus

Se diseñó una red neuronal LSTM bajo un esquema secuencial, en la que se integraron dos capas LSTM de 25 neuronas cada una, configuradas con función de activación ReLU y parámetros `return_sequences` diferenciados para permitir primero la transmisión de secuencias completas y luego la reducción a una salida final. Posteriormente, se añadieron capas densas (13 y 1 neurona) para refinar la representación intermedia y generar la predicción final del diagnóstico (0 = no diabetes, 1 = prediabetes, 2 = diabetes). El modelo se compiló con el optimizador Adam y la función de pérdida MSE, garantizando estabilidad numérica en el entrenamiento. La fase de ajuste se desarrolló en 600 épocas, con batch size de 32 y una validación interna del 10 %.

El aporte directo de este procedimiento fue dotar al modelo de la capacidad de capturar dependencias complejas entre variables clínicas escaladas (edad, IMC, presión arterial, glicemia, HbA1c, tolerancia oral a la glucosa), permitiendo mejorar la precisión en la clasificación de los tres estados de salud. De este modo, la red LSTM no solo entrenó de forma robusta y generalizable, sino que proporcionó una base técnica sólida para consolidar un sistema predictivo confiable y replicable en el contexto clínico de Jaén.

Figura 5. Evolución de la función de pérdida durante el entrenamiento y validación de la red neuronal



Entrenamiento de la red neuronal artificial, en la Figura 11, el eje horizontal representa el número de épocas (600), mientras que el eje vertical representa el error (pérdida) que comete la red neuronal al predecir, tanto en entrenamiento como en validación. La curva azul (loss) corresponde al error de entrenamiento en el conjunto de datos usados para ajustar los pesos de la red, y la curva naranja (val_loss) representa el error de validación en un conjunto de datos distinto.

Durante las primeras etapas del entrenamiento, ambas curvas muestran una caída fuerte. Este es un comportamiento esperado en redes neuronales debido al ajuste inicial de los pesos y evidencia que el modelo aprende rápidamente a reducir sus errores. A medida que avanza el entrenamiento, la pendiente de ambas curvas comienza a disminuir, mostrando una estabilización progresiva a partir de aproximadamente de la época 150.

Un aspecto clave a destacar es que las curvas de pérdida para entrenamiento y validación se mantienen estrecha cercanía entre si a lo largo del proceso; esto sugiere que el modelo no está incurriendo en sobreajustes (overfitting), es decir, no estaba memorizando los datos de entrenamiento, sino que está logrando generalizar adecuadamente sobre datos no vistos. Además, no se observan incrementos abruptos o divergencias en la curva de val_loss. También indica estabilidad en el proceso de aprendizaje.

El valor final de la pérdida, tanto en entrenamiento como en validación, terminan rozando el cero. Lo que refleja que el modelo ha logrado minimizar eficazmente el error en sus predicciones.

Este comportamiento sugiere que el modelo LSTM ha sido entrenado de manera adecuada y ha alcanzado una buena capacidad de generalización, lo cual es fundamental para su posterior aplicación en tareas de predicción con nuevos datos.

A. Bondades del Modelo

- Capacidad de memoria a largo plazo: Ideal para relaciones temporales complejas entre datos (por ejemplo, historial clínico).
- Buena adaptación a datos no lineales: Las redes LSTM con activación ReLU pueden modelar relaciones complejas entre múltiples factores de riesgo.
- Evita el problema del desvanecimiento del gradiente: gracias a su arquitectura interna con "puertas de entrada, olvido y salida".

- Alta precisión en predicciones clínicas: útil cuando se dispone de suficientes datos históricos.
- Modularidad: se puede ajustar el número de neuronas, capas y funciones de activación.
- Uso de validación interna: permite evitar el sobreajuste (overfitting) y mejorar la generalización.
- Robusto ante datos secuenciales: como series temporales de mediciones biomédicas.
- Aplicación en salud predictiva: ayuda en decisiones clínicas tempranas y preventivas.

B. ¿Cómo "aprende" el modelo?

Cuando ejecutas: Durante la fase de entrenamiento, al ejecutar `Red_LSTM.fit(X_train, Y_train, epochs=600, batch_size=32, validation_split=0.1)`, la red neuronal LSTM procesó los datos clínicos de entrada (Edad, IMC, presión arterial, glicemia basal, HbA1c, tolerancia oral a la glucosa) contenidos en `X_train`, junto con las etiquetas de salida en `Y_train` correspondientes al diagnóstico (0 = No diabetes, 1 = Prediabetes, 2 = Diabetes). En este proceso, la red ajustó progresivamente sus pesos internos a lo largo de 600 ciclos de entrenamiento, minimizando el error cuadrático medio entre la predicción generada y el valor real esperado.

El aporte de este procedimiento radica en que el modelo adquirió la capacidad de aprender patrones clínicos relevantes y diferenciarlos con mayor precisión, lo que incrementa su potencial predictivo frente a nuevos pacientes. Esto constituye un paso crucial en la investigación, ya que valida la pertinencia de aplicar LSTM en el contexto de Jaén como herramienta de apoyo para la detección temprana y clasificación del riesgo de diabetes mellitus.

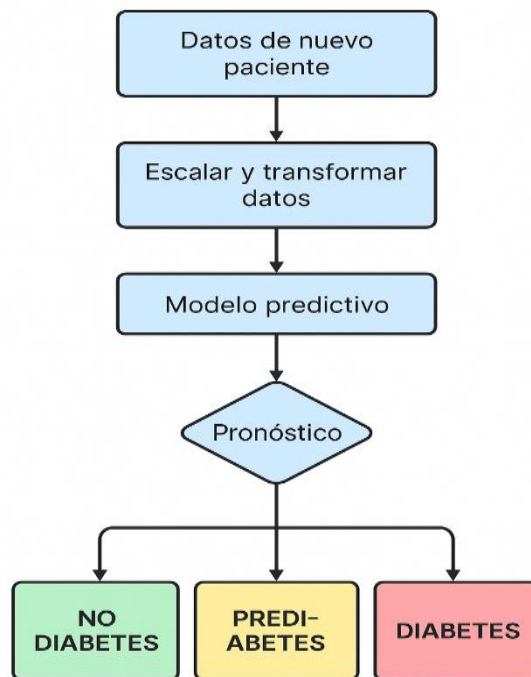
C. ¿Cómo "pronostica" o predice?

Una vez entrenada la red neuronal LSTM, se procedió a su uso en la predicción de nuevos casos clínicos. En esta etapa del análisis, fue indispensable replicar las condiciones iniciales de la red. Por ello, al utilizar los datos de un paciente (ejemplo: Edad = 55 años, IMC = 29.1, presión sistólica = 145 mmHg, entre otros), escalamos y reestructuramos cada uno de estos valores siguiendo idénticos parámetros a los del

entrenamiento. Así logramos que esta nueva información clínica mantuviera una compatibilidad total con los formatos de entrada que nuestro modelo ya reconocía. Al ejecutar la función de predicción, la red entregó una probabilidad asociada al diagnóstico, diferenciando entre los tres posibles estados: 0 = No diabetes, 1 = Prediabetes y 2 = Diabetes.

El aporte principal de este procedimiento es que demuestra la aplicabilidad práctica del modelo entrenado, permitiendo generar predicciones individualizadas a partir de variables clínicas reales. De esta manera, la investigación aporta una herramienta con potencial para apoyar la detección temprana y clasificación del riesgo de diabetes mellitus en Jaén, contribuyendo tanto al ámbito científico como a la práctica médica local.

Figura 6. Diagrama de Flujo del proceso de predicción del estado de Diabetes a partir de un nuevo paciente mediante un modelo predictivo



D. ¿Qué significa el resultado?

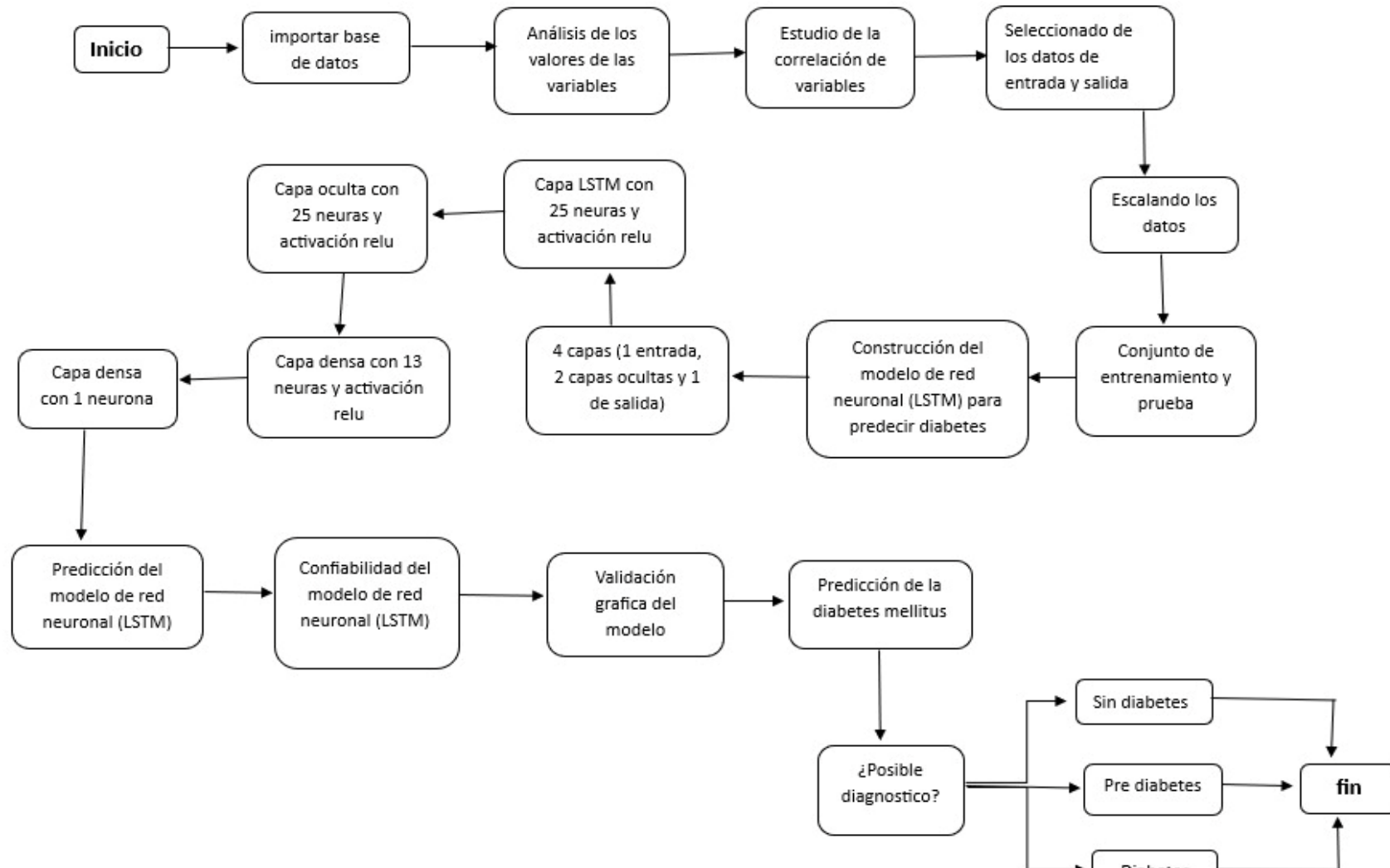
Durante el entrenamiento de la red neuronal LSTM, ejecutado en 600 épocas, el modelo procesó las variables clínicas de entrada (edad, sexo, IMC, presión arterial

sistólica, diastólica y media, glicemia basal, HbA1c y tolerancia oral a la glucosa), ajustando sus parámetros internos mediante el optimizador Adam y la función de pérdida mean squared error. La red generó salidas en un rango continuo [0–2], correspondientes a las categorías: 0 = sin diabetes, 1 = prediabetes y 2 = diabetes, lo que permitió identificar no solo la presencia o ausencia de la enfermedad, sino también un estado intermedio de riesgo metabólico. Este resultado aportó un valor técnico significativo, ya que superó los esquemas de clasificación binaria convencionales, incorporando la detección temprana de prediabetes como soporte para el diagnóstico clínico y la toma de decisiones en salud pública en Jaén – 2023.

E. Consideraciones importantes:

Una vez entrenada la red LSTM, se ingresaron al modelo los valores clínicos reales de un paciente (edad, IMC, presión arterial, glicemia basal, HbA1c, entre otros), los cuales fueron previamente escalados siguiendo el mismo procedimiento aplicado en la etapa de entrenamiento para garantizar la coherencia de los datos. Al procesar esta información, la red generó un valor de salida continuo en el rango de 0 a 2, donde los valores próximos a 0 correspondieron a ausencia de diabetes, aquellos cercanos a 1 indicaron un estado de prediabetes, y los próximos a 2 reflejaron un diagnóstico positivo de diabetes mellitus.

Figura 7. Funcionamiento lógico de la red neuronal artificial



Predicciones del modelo de red neuronal LSTM

La aplicación del método predictivo permitió utilizar la red neuronal LSTM entrenada para estimar la variable de salida Diagnóstico a partir del conjunto de prueba (X_{test}). El modelo, entrenado en 600 épocas con un tamaño de lote de 32 y validación interna del 10%, generó predicciones continuas en el rango [0–2], correspondientes a 0 = sin diabetes, 1 = prediabetes y 2 = diabetes.

El análisis del historial de entrenamiento, representado gráficamente mediante la evolución de la función de pérdida, evidenció la convergencia progresiva del modelo y su capacidad de generalización sobre los datos de validación. Este resultado confirmó la viabilidad técnica de la propuesta, al demostrar que la red LSTM no solo aprendió patrones clínicos relevantes, sino que además pudo transferir ese aprendizaje a nuevos casos, constituyéndose en un aporte innovador para el diagnóstico temprano de la diabetes mellitus en la ciudad de Jaén – 2023.

Evaluación del modelo de red neuronal artificial

Una vez construido el modelo, se debe evaluar su desempeño mediante métricas que indiquen su capacidad de clasificación, tales como: Accuracy (exactitud), Precisión (precisión), Recall (sensibilidad) y F1-Score.

- Accuracy (exactitud)

El modelo alcanzó una exactitud de 0.86746987, lo que significa que acertó aproximadamente el 87% de todas las predicciones realizadas. En términos prácticos, este valor refleja que el modelo tiene un buen desempeño general al clasificar los casos en las tres categorías: no diabetes, prediabetes y diabetes.

- Precisión (precisión)

El modelo alcanzó una precisión de 0.87161601, lo que significa que, de todas las predicciones que el modelo clasificó como positivas para cada clase, aproximadamente el 87% fueron correctas. En términos prácticos, esto indica que el modelo genera pocos falsos positivos y clasifica de manera adecuada los casos asignados a no diabetes, prediabetes y diabetes.

- Recall (sensibilidad)

El modelo alcanzó una sensibilidad de 0.86746987, lo que significa que identificó correctamente alrededor del 87% de los casos reales en las tres categorías de diagnóstico. En términos prácticos, esto indica que el modelo comete pocos falsos negativos y tiene una buena capacidad para detectar correctamente los casos reales de no diabetes, prediabetes y diabetes.

- F1-Score

El modelo obtuvo un F1-Score de 0.86547524, lo que indica que mantiene un buen equilibrio entre precisión y sensibilidad, logrando clasificar correctamente alrededor del 86% de los casos considerando tanto los aciertos como los errores cometidos. En términos prácticos, este resultado muestra que el modelo presenta un desempeño sólido al identificar correctamente los casos de no diabetes, prediabetes y diabetes.

Figura 8. Código en Python para calcular las métricas de evaluación

```
[69]
✓ 0 s
▶ # Accuracy (exactitud), Precision (precisión), Recall (sensibilidad) y F1-Score.
  from sklearn.metrics import accuracy_score
  print("Accuracy (Exactitud) :", accuracy_score(Diagnóstico_real, Pred_clases))
  from sklearn.metrics import precision_score
  print("Precision (Precisión) :", precision_score(Diagnóstico_real, Pred_clases, average='weighted'))
  from sklearn.metrics import recall_score
  print("Recall (Sensibilidad) :", recall_score(Diagnóstico_real, Pred_clases, average='weighted'))
  from sklearn.metrics import f1_score
  print("F1-Score :", f1_score(Diagnóstico_real, Pred_clases, average='weighted'))

▼ ... Accuracy (Exactitud) : 0.8674698795180723
    Precision (Precisión) : 0.8716160187307935
    Recall (Sensibilidad) : 0.8674698795180723
    F1-Score : 0.8654752441939738
```

Validación gráfica del modelo

La validación gráfica del modelo se ejecutó como complemento al análisis estadístico, con el fin de evaluar visualmente la concordancia entre los valores reales del diagnóstico y las predicciones obtenidas mediante la red neuronal LSTM. Para ello, tanto los datos de salida reales (Y_test) como las predicciones del modelo fueron reescalados a su rango original con el MinMaxScaler, lo que permitió construir una tabla comparativa

con tres elementos: Diagnóstico_real, Predicción_LSTM y la diferencia absoluta entre ambos.

Posteriormente, se graficaron las series de valores reales y predichos, obteniéndose curvas con alta superposición, lo cual evidenció que el modelo reprodujo de manera consistente las tendencias clínicas observadas en la base de datos.

Desde el punto de vista técnico, el proceso de validación nos permitió confirmar que el algoritmo logró capturar el comportamiento de las variables clínicas. Más importante aún, las salidas que arrojó el sistema mantuvieron una estricta coherencia con las tres categorías de diagnóstico establecidas (0 = sin diabetes, 1 = prediabetes y 2 = diabetes). Para los fines de esta investigación, obtener este resultado se traduce en una garantía extra de confiabilidad. Nos demuestra que la propuesta va mucho más allá de simplemente lograr buenas métricas cuantitativas; por ello se evidencia una clara correspondencia gráfica entre lo que el algoritmo predice y los diagnósticos reales de los pacientes. En definitiva, todo este proceso fortalece de manera significativa la pertinencia de utilizar este modelo para la detección temprana de diabetes mellitus en Jaén – 2023.”

Tabla 10. Registro de datos de entrada del Hospital – San Javier de Bellavista, ciudad de Jaén

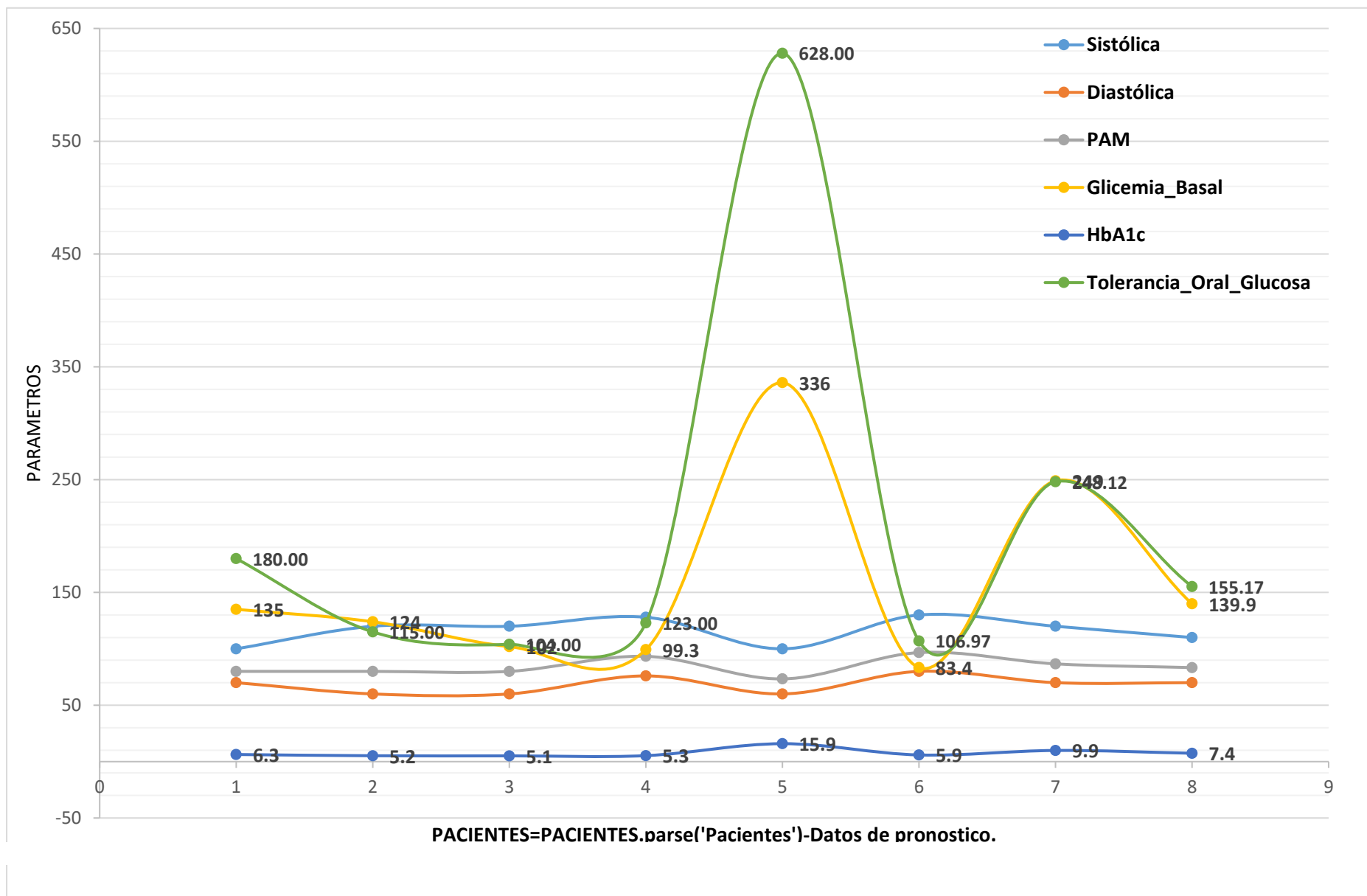
Edad	Sexo	IMC	Sistólica	Diastólica	PAP	Glicemia Basal	HbA1c	Tolerancia_Oral Glucosa
51	F	25.3	100	70	80.00	135	6.3	180.00
61	F	22.6	120	60	80.00	124	5.2	115.00
62	F	22.6	120	60	80.00	102	5.1	104.00
65	M	38	128	76	93.33	99.3	5.3	123.00
55	F	28.2	100	60	73.33	336	15.9	628.00
62	F	25.4	130	80	96.67	83.4	5.9	106.97
61	F	28	120	70	86.67	249	9.9	248.12
61	F	25.5	110	70	83.33	139.9	7.4	155.17

Descripción e interpretación: Al analizar los datos de la segunda muestra, correspondiente a pacientes entre 51 y 65 años según la Tabla, se notan diferencias claras en los indicadores clínicos vinculados al metabolismo de la glucosa. Al observar los niveles de glicemia en ayunas, hemoglobina glicosilada (HbA1c) y los resultados de la prueba de tolerancia oral a la glucosa, se identifican con facilidad casos que reflejan una desregulación metabólica evidente.

Dos pacientes destacan por presentar valores considerablemente elevados. El caso más preocupante es el de una mujer de 55 años, cuya glicemia basal alcanza los 336 mg/dL, con una HbA1c de 15.9 % y una prueba de tolerancia oral que llega a 628 mg/dL. Estas cifras muestran un cuadro avanzado de hiperglucemia crónica y un serio descontrol glucémico. De manera similar, encontramos un paciente de 61 años presenta una glicemia en ayunas de 249 mg/dL, la HbA1c de 9.9 % y tolerancia oral de 248.12 mg/dL, lo que también apunta a una diabetes no controlada.

En el nivel intermedio se identificó una paciente de 51 años con glicemia basal de 135 mg/dL, HbA1c de 6.3 % y tolerancia oral a la glucosa de 180 mg/dL, parámetros compatibles con un estado de riesgo metabólico (prediabetes). En contraste, los demás pacientes presentaron valores dentro de rangos normales: glicemia basal <125 mg/dL, HbA1c <6.0 % y tolerancia oral <140 mg/dL, evidenciando un adecuado control glucémico.

Figura 9. Estadígrafo del registro de datos pronósticos del Hospital San Javier de Bellavista, ciudad de Jaén



Se puede afirmar que el modelo muestra una capacidad adecuada para detectar casos compatibles con diabetes mellitus, al identificar con precisión patrones alterados en variables clave como glicemia basal, hemoglobina glicosilada (HbA1c) y la prueba de tolerancia oral a la glucosa.

En el caso de los datos reales, los pacientes sin diagnóstico (diagnóstico = 0) muestran parámetros clínicos dentro de los rangos de referencia:

La ADA (Asociación Americana de Diabetes), considera estos valores de referencia para **Glicemia en ayunas (Fasting Plasma Glucose, FPG)** ⁽⁴⁷⁾:

- Normal: < 100 mg/dL
- Prediabetes (glucosa en ayunas alterada): 100 – 125 mg/dL
- Diabetes: \geq 126 mg/dL

Además, considera los siguientes valores para la **Hemoglobina glicosilada (HbA1c)** ⁽⁴⁷⁾:

- Normal: < 5.7 %
- Prediabetes: 5.7 % – 6.4 %
- Diabetes: \geq 6.5 %

Mientras que toma en cuenta los siguientes valores para el **Test de tolerancia oral a la glucosa (OGTT)** ⁽⁴⁷⁾:

- Normal: < 140 mg/dL
- Prediabetes (tolerancia alterada): 140 – 199 mg/dL
- Diabetes: \geq 200 mg/dL

Estos resultados indican un control glucémico dentro de los rangos normales, incluso en algunos pacientes con sobrepeso u obesidad. En cambio, los pacientes con diagnóstico 1 o 2 presentan:

- Glicemia en ayunas \geq 126 mg/dL
- HbA1c \geq 6.5 %
- Tolerancia oral \geq 200 mg/Dl

Lo que valida la relación entre alteraciones clínicas y el diagnóstico médico. Ahora, al analizar la salida del modelo de pronóstico, se observa que el sistema logra identificar correctamente los perfiles críticos:

- El caso más severo predicho es el de una paciente con glicemia de 336 mg/dL, HbA1c de 15.9 % y tolerancia oral de 628 mg/dL, valores consistentemente compatibles con diabetes mellitus tipo 2 descontrolada.
- Otro caso con 249 mg/dL de glicemia y 9.9 % de HbA1c también fue identificado como diagnóstico positivo, lo que coincide con los parámetros clínicos típicos de diabetes en fase activa.

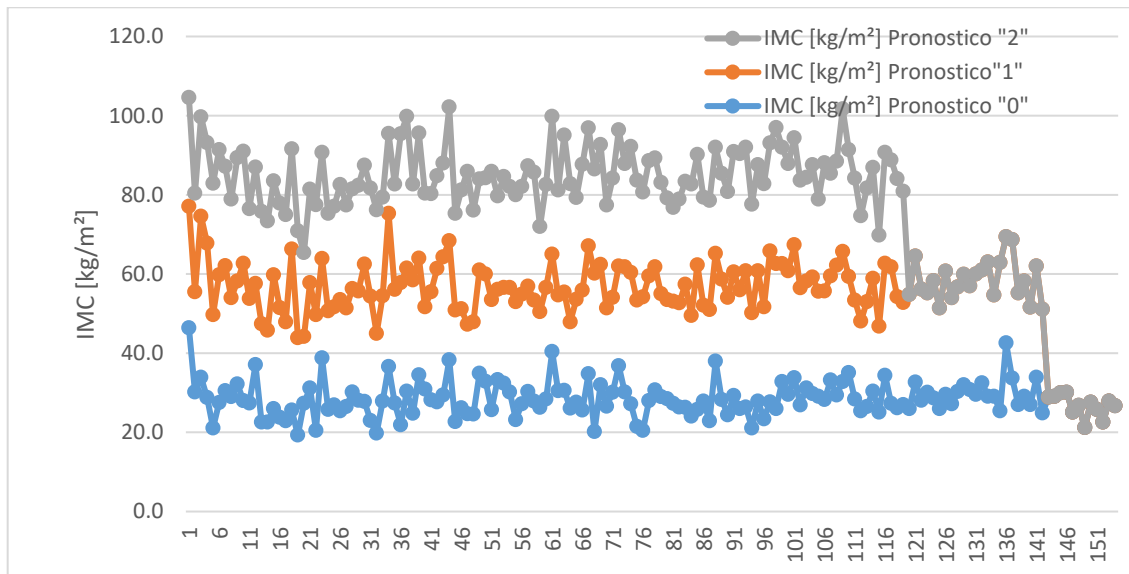
Además, el modelo muestra sensibilidad al detectar pacientes en condición de riesgo o prediabetes, como en el caso de una paciente con 135 mg/dL de glicemia, 6.3 % de HbA1c y 180 mg/dL en la prueba de tolerancia, valores que superan los umbrales normales, pero no alcanzan niveles críticos, lo que permite alertar sobre un posible desarrollo futuro de la enfermedad.

El modelo de red neuronal artificial ha demostrado una buena capacidad predictiva al reflejar patrones similares a los observados clínicamente:

- Identifica correctamente los casos severos de diabetes con múltiples variables alteradas.
- Diferencia adecuadamente a los pacientes con control glucémico de aquellos con riesgo o alteraciones.
- Se alinea con los rangos clínicos definidos por estándares médicos internacionales, lo cual valida su utilidad diagnóstica.

El modelo LSTM demostró su capacidad para predecir con precisión los casos de hiperglucemia severa, niveles elevados de HbA1c y tolerancia oral a la glucosa, así como para identificar perfiles intermedios compatibles con prediabetes. Esto evidencia su potencial como herramienta de apoyo diagnóstico en la práctica clínica, reforzando su validez técnica para la detección temprana de diabetes mellitus en la ciudad de Jaén – 2023. Para su aplicación en pacientes externos a la base de entrenamiento, los datos clínicos fueron escalados mediante el mismo procedimiento de normalización (MinMaxScaler) utilizado en la etapa de entrenamiento, asegurando la coherencia en la interpretación de las variables. Posteriormente, se aplicó la red neuronal entrenada, generando predicciones continuas que fueron ajustadas al rango de clasificación definido (0 = sin diabetes, 1 = prediabetes, 2 = diabetes), lo que permitió evaluar el estado de cada paciente con base en sus características clínicas individuales.

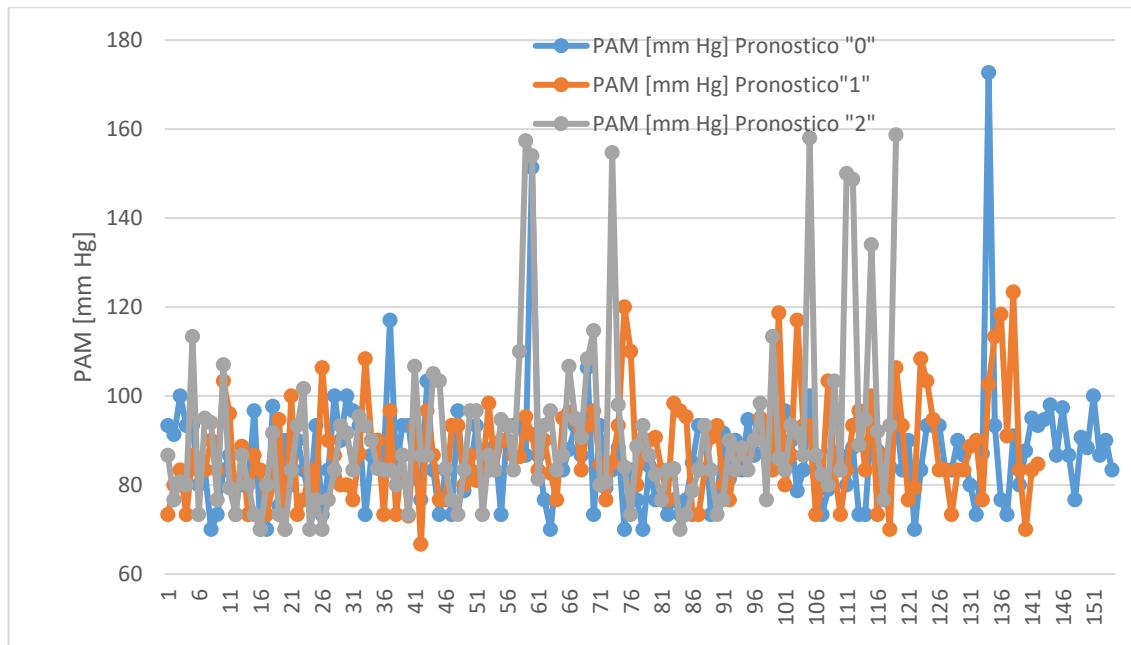
Figura 10. Diagrama de dispersión de valores de IMC



Descripción e interpretación: El gráfico presentado muestra la relación entre el Índice de Masa Corporal (IMC) y el pronóstico clínico de los pacientes, clasificados en tres niveles: pronóstico leve ("0"), moderado ("1") y severo ("2"). Se observa una tendencia clara: a mayor severidad del pronóstico, mayor es el IMC registrado. Los pacientes con pronóstico "2" concentran los valores más altos de IMC, alcanzando incluso valores mayores a 100 kg/m², lo que sugiere una presencia significativa de obesidad mórbida en este grupo. En cambio, el grupo con pronóstico "0" presenta valores considerablemente más bajos, oscilando mayoritariamente entre los 20 y 40 kg/m². Esta diferencia sugiere que existe una posible asociación directa entre el exceso de peso corporal y el agravamiento del estado clínico.

Asimismo, se evidencia una mayor dispersión en los valores de IMC en los pacientes con pronóstico severo, lo que podría estar relacionado con la presencia de múltiples comorbilidades. Estos hallazgos respaldan la hipótesis de que un IMC elevado podría constituir un factor de riesgo significativo en la evolución desfavorable de pacientes, lo que concuerda con estudios previos sobre los efectos del sobrepeso y la obesidad en la salud general.

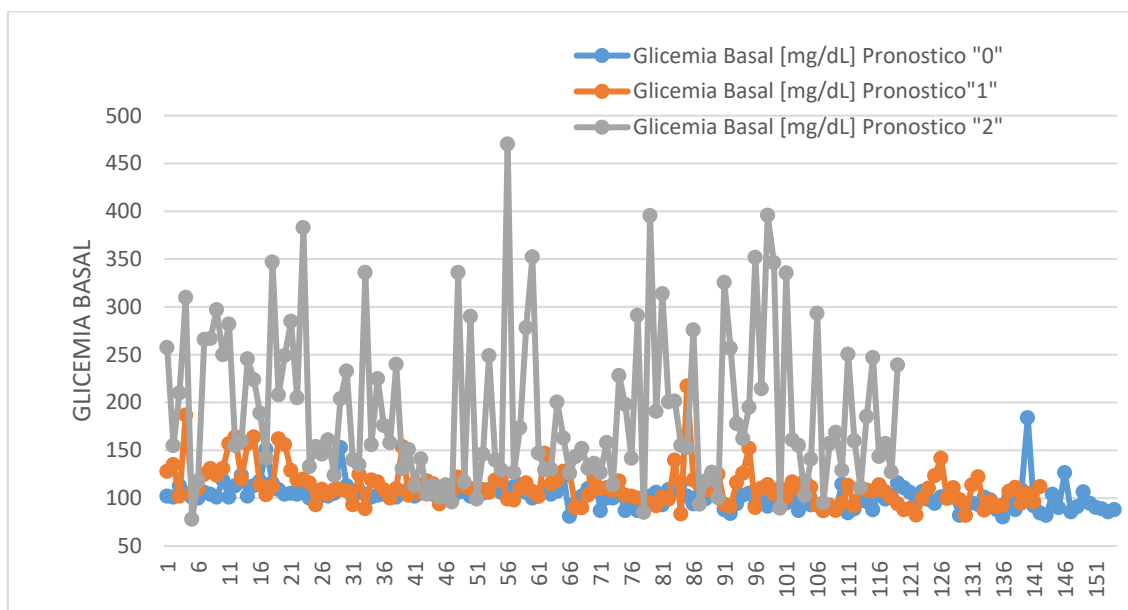
Figura 11. Diagrama de dispersión de valores del PAM



Descripción e interpretación: En la gráfica se observa el comportamiento de la Presión Arterial Media (PAM) clasificada según el pronóstico clínico identificado como “0”, “1” y “2”. Se evidencia que el grupo con pronóstico "0" presenta valores de PAM relativamente estables y dentro de rangos normales, en su mayoría entre 80 y 100 mm Hg, con pocos casos que superan los 120 mm Hg. Por otro lado, el grupo con pronóstico "1" muestra una mayor dispersión, aunque la concentración principal también se mantiene en el rango normotenso, con algunos valores aislados por encima de 120 mm Hg.

Finalmente, el grupo con pronóstico "2" presenta la mayor variabilidad y valores extremos, con múltiples casos que superan los 140 mm Hg e incluso alcanzan niveles cercanos a 160 mm Hg, lo que sugiere hipertensión significativa o desregulación cardiovascular. Esta distribución sugiere que a mayor nivel de pronóstico (indicando mayor riesgo o severidad), mayor es la alteración en los valores de presión arterial media, lo cual podría relacionarse con complicaciones metabólicas o cardiovasculares en el grupo evaluado.

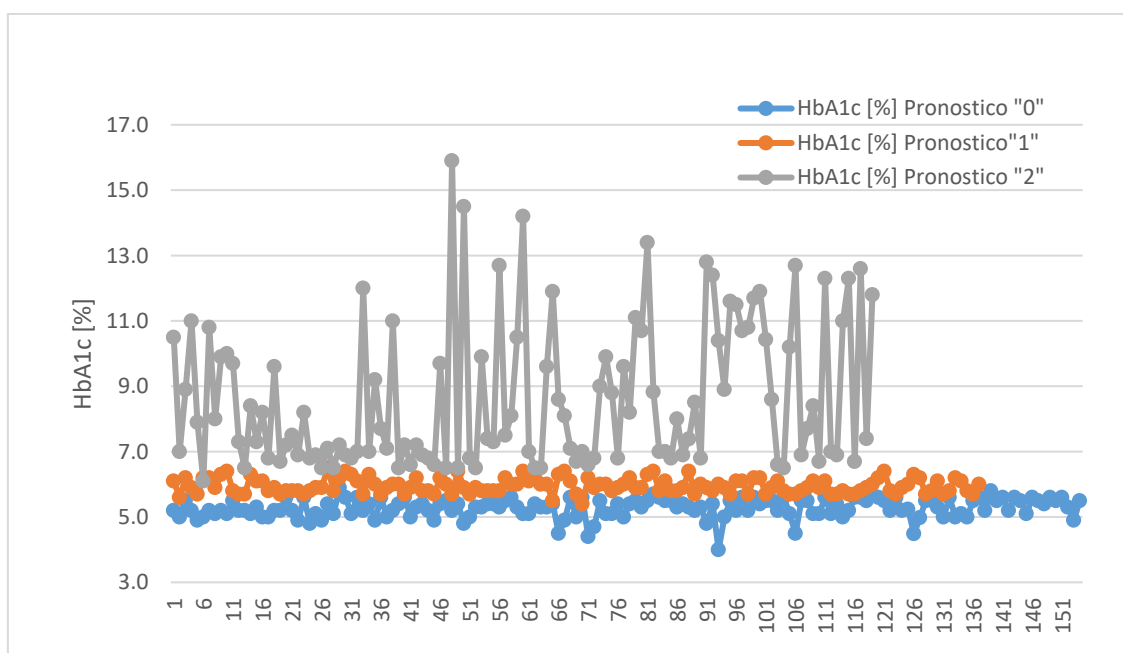
Figura 12. Diagrama de dispersión de valores de la Glicemia



Descripción e interpretación: La Figura, presenta la distribución de los valores de glicemia basal (mg/dL) diferenciados por pronóstico clínico: “0”, “1” y “2”. Se observa que los individuos con pronóstico "0" (curva azul) presentan valores de glicemia mayoritariamente dentro del rango normal o levemente elevados, oscilando entre 90 y 120 mg/dL, con pocos valores que superan los 140 mg/dL. En contraste, los valores correspondientes al pronóstico "1" (curva naranja) muestran una tendencia ascendente moderada, con una mayor concentración de datos entre 120 y 160 mg/dL, lo cual se relaciona con condiciones de prediabetes o alteración de la glucosa en ayunas.

No obstante, es en el grupo con pronóstico "2" (curva gris) donde se evidencia la mayor dispersión y presencia de valores extremos, alcanzando niveles superiores a 400 mg/dL, lo que refleja casos de hiperglicemia severa y descontrol glucémico. Esta tendencia sugiere que, a mayor nivel de pronóstico, los valores de glicemia basal se incrementan significativamente, indicando una posible asociación entre niveles elevados de glucosa en ayunas y un mayor riesgo clínico o presencia de diabetes mellitus no controlada. En resumen, la glicemia basal demuestra ser un marcador sensible para diferenciar estados de salud metabólica, apoyando su uso como variable relevante en modelos predictivos de riesgo.

Figura 13. Diagrama de dispersión de valores HbA1c [%]

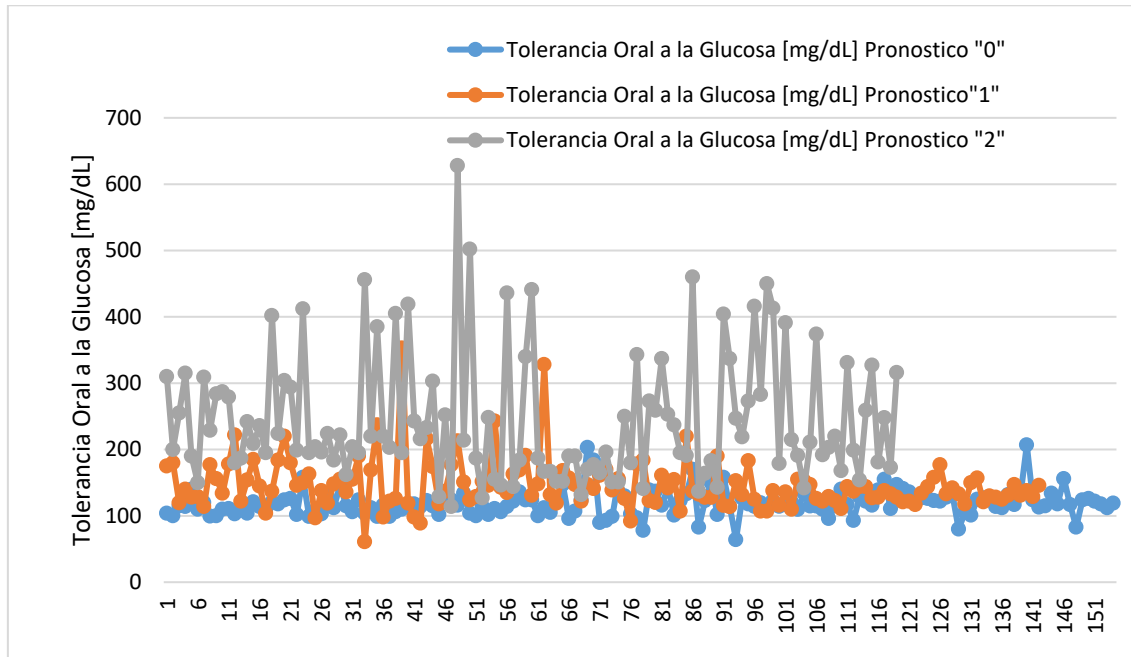


Descripción e interpretación: La Figura, visualiza los niveles de hemoglobina glucosilada (HbA1c) expresados en porcentaje, agrupados según el pronóstico clínico: “0”, “1” y “2”. Se observa que el grupo con pronóstico "0" (línea azul) presenta valores de HbA1c consistentemente bajos, en su mayoría entre 5,0% y 5,7%, lo que se asocia con un adecuado control glucémico o la ausencia de diabetes. En el grupo con pronóstico "1" (línea naranja), los valores de HbA1c tienden a situarse en un rango ligeramente superior, entre 5,7% y 6,4%, lo que corresponde a un estado de prediabetes o tolerancia alterada a la glucosa, según los criterios diagnósticos internacionales.

En cambio, el grupo con pronóstico "2" (línea gris) muestra una amplia dispersión y elevación considerable de los valores de HbA1c, con múltiples casos que superan el 7,0%, y valores extremos por encima de 13% e incluso 15%, lo cual indica diabetes mal controlada o crónica. Esta tendencia refleja que, a mayor severidad del pronóstico clínico, se incrementa significativamente el valor de HbA1c, posicionándolo como un indicador clave del deterioro metabólico y del riesgo asociado.

Por tanto, el análisis de la HbA1c permite no solo clasificar el estado glucémico de los pacientes, sino también inferir su riesgo clínico a futuro, constituyéndose en una variable relevante para modelos predictivos y sistemas de alerta temprana.

Figura 14. Diagrama de dispersión de valores de la Tolerancia Oral a la Glucosa [mg/dL]



Descripción e interpretación: Se muestra, los niveles de Tolerancia Oral a la Glucosa [mg/dL], expresados en porcentaje, agrupados según el pronóstico clínico: “0”, “1” y “2”. Se observa que el grupo con pronóstico “0” (línea azul) presenta valores de [mg/dL], predominantemente menores a 140 mg/dL, lo que se asocia con un adecuado Tolerancia Oral a la Glucosa [mg/dL] y la ausencia de diabetes. En el grupo con pronóstico “1” (línea naranja), los valores tienden a situarse en un rango entre 140 y 199 mg/dL, lo que corresponde a un estado de prediabetes o tolerancia alterada a la glucosa, según los criterios diagnósticos internacionales.

En cambio, el grupo con pronóstico “2” (línea gris) muestra una amplia dispersión y elevación considerable de los valores de c con múltiples casos que superan valores extremos por encima de 150 a 600, lo cual indica diabetes mal controlada o crónica. Por tanto, el análisis de la [mg/dL], permitió no solo clasificar el estado glucémico de los pacientes, sino también inferir su riesgo clínico a futuro, constituyéndose en una variable relevante para modelos predictivos y sistemas de alerta temprana.

Resultados de las predicciones

Tabla 11. Resultados de la predicción de diabetes mellitus

Modelo	Resultado
0.6579536	1= prediabetes
-0.4149455	0 = sin diabetes
-0.5233955	1 = prediabetes
-0.99918354	1 = prediabetes
2.4623919	2 = diabetes
-0.32902876	0 = sin diabetes
1.9825699	2 = diabetes
1.5005553	2 = diabetes
...

Interpretación: El modelo de red neuronal generó salidas numéricas que, tras su interpretación y clasificación, permitieron diferenciar tres estados clínicos: sin diabetes (0), prediabetes (1) y diabetes (2). En la Tabla se muestra un extracto de los resultados obtenidos. Por ejemplo, un valor de 0.6579 fue clasificado como prediabetes (1), mientras que valores negativos como -0.4149 y -0.3290 fueron asociados a ausencia de la enfermedad (0). Por otro lado, valores elevados como 2.4623, 1.9825 y 1.5005 correspondieron a la categoría diabetes (2). Estos resultados evidenciaron que el modelo propuesto logró discriminar entre individuos sanos, en riesgo y con diagnóstico positivo de diabetes mellitus, validando su utilidad como herramienta de predicción clínica en la ciudad de Jaén.

ANEXO 3: SOLICITUD Y AUTORIZACIÓN

DIRECCION DE CAPACITACION
"AÑO DEL BICENTENARIO, DE LA CONSOLIDACION DE NUESTRA INDEPENDENCIA, Y DE LA
CONMEMORACION DE LAS HEROICAS BATALLAS DE JUNIN Y AYACUCHO"

Jaén, 03 de Junio del 2024

MAD 09598431

OFICIO N° 017 - 2024- GR.CAJ/DSRSJ-DG/DEGDRRH-CAPACITACION

MED. CIRUJ. SAUL CORDOVA CHAPOÑAN
DIRECTOR DE HOSPITAL "SAN JAVIER DE BELLAVISTA"

PRESENTE.-



Firmado digitalmente por LEON
SANCHEZ Dwaygh Manuel FAU
20453744183 soft
Motivo: Visto en señal de
conformidad
Fecha: 03/06/2024 12:10 p. m.

**ASUNTO: AUTORIZACION PARA EJECUCION DE PROYECTO
DE INVESTIGACION**
REF. OFICIO N°001-24-REG-CAJ/SRS-J/HASJB/RC

Es sumamente grato dirigirme a usted con la finalidad de hacerle llegar el saludo cordial y al mismo tiempo hacerle de su conocimiento que esta Dirección autoriza la ejecución del proyecto de investigación en los centros de salud de la jurisdicción de la Red Integrada de Salud Jaén.

1. Proyecto de Investigación : "Inteligencia artificial para diagnosticar diabetes mellitus en Jaén - Perú"
2. Responsables : Bach. Clara Cerna Reátegui y Bach. Anyelí Molocho Martínez.
3. Universidad : Universidad Nacional de Jaén.

Asimismo, mencionar a usted, que deberá alcanzar la Declaración Jurada de Confidencialidad del manejo de datos, así como un ejemplar de la tesis desarrollada.

Es propicia la oportunidad para expresarle las muestras de consideración.

Atentamente,

EHMV/DMLS.
C.c.
- Archivo

GOBIERNO REGIONAL CAJAMARCA
DIRECCION REGIONAL DE SALUD
M.C. Edward Henry Mundaca Vidarte
DIRECTOR GENERAL

ANEXO 5: PROCESO DE VALIDACIÓN DEL INSTRUMENTO DE RECOLECCIÓN DE DATOS





CONSTANCIA DE VALIDACIÓN DE INSTRUMENTOS

Quien suscribe, José Guillermo Samamé Céspedes con documento de identidad N° 16719728, de profesión _____, ejerciendo actualmente como docente, en la Universidad Nacional de Jaén. Por medio del presente hago constar que he revisado con fines de Validación el Instrumento (ficha recolectora de datos) a los efectos de su aplicación en el informe de tesis con título: **INTELIGENCIA ARTIFICIAL PARA DIAGNOSTICAR DIABETES MELLITUS EN JAEN-PERU 2023.**, de las autoras Br. Clara Cerna Reátegui y Br. Evelyn Anyell Molocho Martínez.

Luego de hacer las observaciones pertinentes, puedo formular las siguientes apreciaciones.

	Deficiente	Aceptable	Bueno	Excelente
Coherencia de ítems			X	
Amplitud de contenido			X	
Redacción de ítems			X	
Claridad			X	
Precisión			X	

Fecha: Jaén, 03 de Enero del 2024

 UNIVERSIDAD NACIONAL DE JAÉN


Dr. José Guillermo Samamé Céspedes
PROFESOR PRINCIPAL TIEMPO PARCIAL
Escuela Profesional Tecnología Médica

Firma

DNI: 16719728



CONSTANCIA DE VALIDACIÓN DE INSTRUMENTOS

Quien suscribe, Angel Wilmer Paredes Guerrero con documento de identidad N° 08141134, de profesión Ing de Sistemas, ejerciendo actualmente como docente, en la Universidad Nacional de Jaén. Por medio del presente hago constar que he revisado con fines de Validación el Instrumento (ficha recolectora de datos) a los efectos de su aplicación en el informe de tesis con título: **INTELIGENCIA ARTIFICIAL PARA DIAGNOSTICAR DIABETES MELLITUS EN JAEN-PERU 2023.**, de las autoras Br. Clara Cerna Reátegui y Br. Evelyn Anyelí Molocho Martínez.

Luego de hacer las observaciones pertinentes, puedo formular las siguientes apreciaciones.

	Deficiente	Aceptable	Bueno	Excelente
Coherencia de ítems				X
Amplitud de contenido			X	
Redacción de ítems				X
Claridad				X
Precisión				X

Fecha: Jaén, 2 de enero del 2024

Firma

DNI: 08141134




CONSTANCIA DE VALIDACIÓN DE INSTRUMENTOS

Quien suscribe, Adán Díaz Ruiz con documento de identidad N° 10776471, de profesión Químico, ejerciendo actualmente como docente, en la Universidad Nacional de Jaén. Por medio del presente hago constar que he revisado con fines de Validación el Instrumento (ficha recolectora de datos) a los efectos de su aplicación en el informe de tesis con título: **INTELIGENCIA ARTIFICIAL PARA DIAGNOSTICAR DIABETES MELLITUS EN JAEN-PERU 2023.**, de las autoras Br. Clara Cerna Reátegui y Br. Evelyn Anyell Molocho Martínez.

Luego de hacer las observaciones pertinentes, puedo formular las siguientes apreciaciones.

	Deficiente	Aceptable	Bueno	Excelente
Coherencia de ítems				X
Amplitud de contenido				X
Redacción de ítems				X
Claridad				X
Precisión				X

Fecha: Jaén, 29 de Diciembre del 2023


Firma
DNI: 10776471

ANEXO 6: COMPROMISO DE ASESORIA

UNIVERSIDAD NACIONAL DE JAÉN

Ley de Creación N° 29304

Universidad Licenciada con Resolución del Consejo Directivo N° 002-2018

SUNEDU/CD

FORMATO 01: COMPROMISO DEL ASESOR

El que suscribe, Nilton Arei Fernández, con Doctorado
en Ciencias Matemáticas, identificado
con DNI N° 45071336, con conocimiento del Reglamento General de Grado Académico
y Título Profesional de la Universidad Nacional de Jaén, se compromete y deja constancia
de las orientaciones a las Estudiantes Egresadas.

CLARA CERNA REÁTEGUI y **EVELYN ANYELÍ MOLOCHO MARTÍNEZ**, de
la Escuela Profesional de Tecnología Médica, en la formulación y ejecución del:

- Plan de Trabajo de Investigación Informe Final de Trabajo de Investigación
 Proyecto de Tesis Informe Final de Tesis
 Informe Final del Trabajo por Suficiencia Profesional

Por lo indicado doy testimonio y visto bueno que las asesoradas ha ejecutado el trabajo
de investigación; por lo que en fe a la verdad suscribe la presente.

Jaén, 25 de Septiembre del 2025


Asesor

UNIVERSIDAD NACIONAL DE JAÉN

Ley de Creación N° 29304

Universidad Licenciada con Resolución del Consejo Directivo N° 002-2018

SUNEDU/CD

FORMATO 01: COMPROMISO DEL ASESOR

El que suscribe, Guillermo Núñez Sánchez, con Doctorado
en Gestión Pública y Gobernabilidad, identificado
con DNI N° 45071336, con conocimiento del Reglamento General de Grado Académico
y Título Profesional de la Universidad Nacional de Jaén, se compromete y deja constancia
de las orientaciones a las Estudiantes Egresadas.

CLARA CERNA REÁTEGUI y **EVELYN ANYELÍ MOLOCHO MARTÍNEZ**, de
la Escuela Profesional de Tecnología Médica, en la formulación y ejecución del:

- Plan de Trabajo de Investigación Informe Final de Trabajo de Investigación
 Proyecto de Tesis Informe Final de Tesis
 Informe Final del Trabajo por Suficiencia Profesional

Por lo indicado doy testimonio y visto bueno que las asesoradas ha ejecutado el trabajo
de investigación; por lo que en fe a la verdad suscribe la presente.

Jaén, 25 de Septiembre del 2025

 UNIVERSIDAD NACIONAL DE JAÉN
Guillermo Núñez Sánchez
Dr. Guillermo Núñez Sánchez
PROFESOR AJAENO PRIMARIO COMPLETO
ESCUELA PROFESIONAL DE TECNOLOGÍA MÉDICA

Asesor

ANEXO 7: DECLARACION JURADA

ANEXO 6: DECLARACION JURADA

UNIVERSIDAD NACIONAL DE JAÉN

Ley de Creación

Nº 29304

Universidad Licenciada con Resolución del Consejo Directivo Nº 002-2018-SUNEDU/CD

FORMATO 04: DECLARACIÓN JURADA DE NO PLAGIO


Yo, Clara Cerna Reátegui, con DNI Nº 75468866 estudiante de la Escuela Profesional de Tecnología Médica de la Universidad Nacional de Jaén; declaro bajo juramento que Soy Autor del Proyecto de tesis;

“PROPUESTA DE UN MODELO DE RED NEURONAL ARTIFICIAL PARA PREDECIR DIABETES MELLITUS EN LA CIUDAD DE JAÉN-2023.”

1. El mismo que presento para optar: () Grado Académico de Bachiller (X) Título Profesional
2. El **Proyecto de tesis** no ha sido plagiado ni total ni parcialmente, para la cual se han respetado las normas internacionales de citas y referencias para las fuentes consultadas.
3. El **Proyecto de tesis presentado** no atenta contra derechos de terceros.
4. El **Proyecto de tesis** no ha sido publicado ni presentado anteriormente para obtener algún grado académico previo o título profesional.
5. Los datos presentados en los resultados son reales, no han sido falsificados, ni duplicados, ni copiados. Por lo expuesto, mediante la presente asumo toda responsabilidad que pudiera derivarse por la autoría, originalidad y veracidad del contenido del Proyecto de tesis, así como por los derechos sobre la obra 'invención' presentada. Asimismo, por la presente me comprometo a asumir además todas las cargas pecuniarias que pudieran derivarse para la UNJ en favor de terceros por motivo de acciones, reclamaciones o conflictos derivados del incumplimiento de lo declarado o las que encontraren causa en el contenido del Proyecto de tesis

De identificarse fraude, piratería, plagio, falsificación o que el proyecto de tesis haya sido publicado anteriormente; asumo las consecuencias y sanciones civiles y penales que de mi acción se deriven.

Jaén, 25 Septiembre del 2025.


Clara Cerna Reátegui
DNI: 75468866

UNIVERSIDAD NACIONAL DE JAÉN

Ley de Creación

Nº 29304

Universidad Licenciada con Resolución del Consejo Directivo Nº 002-2018-SUNEDU/CD

FORMATO 04: DECLARACIÓN JURADA DE NO PLAGIO

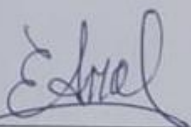
Yo, Evelyn Anyelí Molocho Martínez, con DNI Nº 71869858 estudiante de la Escuela Profesional de Tecnología Médica de la Universidad Nacional de Jaén; declaro bajo juramento que Soy Autor del Proyecto de tesis;

“PROPUESTA DE UN MODELO DE RED NEURONAL ARTIFICIAL PARA PREDECIR DIABETES MELLITUS EN LA CIUDAD DE JAÉN-2023.”

1. El mismo que presento para optar: () Grado Académico de Bachiller (X) Título Profesional
2. El **Proyecto de tesis** no ha sido plagiado ni total ni parcialmente, para la cual se han respetado las normas internacionales de citas y referencias para las fuentes consultadas.
3. El **Proyecto de tesis presentado** no atenta contra derechos de terceros.
4. El **Proyecto de tesis** no ha sido publicado ni presentado anteriormente para obtener algún grado académico previo o título profesional.
5. Los datos presentados en los resultados son reales, no han sido falsificados, ni duplicados, ni copiados. Por lo expuesto, mediante la presente asumo toda responsabilidad que pudiera derivarse por la autoría, originalidad y veracidad del contenido del Proyecto de tesis, así como por los derechos sobre la obra 'invención' presentada. Asimismo, por la presente me comprometo a asumir además todas las cargas pecuniarias que pudieran derivarse para la UNJ en favor de terceros por motivo de acciones, reclamaciones o conflictos derivados del incumplimiento de lo declarado o las que encontraren causa en el contenido del Proyecto de tesis

De identificarse fraude, piratería, plagio, falsificación o que el proyecto de tesis haya sido publicado anteriormente; asumo las consecuencias y sanciones civiles y penales que de mi acción se deriven.

Jaén, 25 Septiembre del 2025.



Evelyn Anyelí Molocho Martínez

DNI: 71869858

ANEXO 9: FOTOGRAFÍAS DE EJECUCIÓN DEL ESTUDIO

Revisión de resultados de laboratorio del hospital san Javier de bellavista.



Revisión de historias clínicas de los pacientes atendidos en el hospital san Javier de bellavista



Elaboración del informe final de tesis

