Revista Multidisciplinaria RIIDG
ISSN: 3091-194X
Doi: https://doi.org/10.64041/riidg.v3i4.30

Revista Internacional de Investigación y Desarrollo Global

Periocidad trimestral, Volumen 3, Numero 4, Años (2024), Pag. 49-67

Modelo predictivo para la detección temprana de diabetes tipo II basado en registros electrónicos de salud

Predictive model for early detection of type II diabetes based on electronic health records

AUTORES

Lianmoy Noemi Facuy Toledo
Universidad Católica de Santiago de Guayaquil
Guayaquil-Ecuador
lianmoy.facuy@cu.ucsg.edu.ec
https://orcid.org/0009-0008-1730-0460

Como citar: Facuy Toledo, L. N. (2024). Modelo predictivo para la detección temprana de diabetes tipo II basado en registros electrónicos de salud. *Revista Internacional De Investigación Y Desarrollo Global*, *3*(4), 49–67. https://doi.org/10.64041/riidg.v3i4.30	Fecha de recepción:2024-10-15 Fecha de aceptación: 2024-11-15 Fecha de publicación:2024-12-15

Revista Multidisciplinaria RIIDG
ISSN: 3091-194X
Doi: https://doi.org/10.64041/riidg.v3i4.30

Revista Internacional de Investigación y Desarrollo Global

Resumen

La diabetes mellitus tipo II es una enfermedad crónica que representa una de las principales causas de morbilidad y mortalidad a nivel mundial, con un incremento sostenido en su prevalencia debido a factores genéticos, ambientales y de estilo de vida. Ante esta problemática, el presente estudio desarrolla un modelo predictivo orientado a la detección temprana de esta patología, a partir del análisis de registros electrónicos de salud (EHR), los cuales contienen información clínica relevante de los pacientes recopilada en entornos hospitalarios y centros de atención primaria.

El modelo se fundamenta en técnicas de aprendizaje automático (machine learning), con énfasis en algoritmos supervisados como regresión logística, máquinas de soporte vectorial, árboles de decisión, random forest y redes neuronales artificiales. El conjunto de datos utilizado fue sometido a un riguroso proceso de limpieza, transformación y selección de características, garantizando la calidad de los datos antes del entrenamiento del modelo. Se consideraron variables clínicas y demográficas como la edad, género, índice de masa corporal (IMC), niveles de glucosa en ayunas, presión arterial, historial familiar de diabetes, perfil lipídico, entre otras.

Los resultados del estudio evidencian que el modelo desarrollado logra una alta capacidad predictiva, con métricas destacadas en precisión, sensibilidad, especificidad y área bajo la curva ROC. Estas métricas permiten validar la utilidad del sistema como herramienta de apoyo para la toma de decisiones médicas, facilitando la identificación de pacientes en riesgo incluso antes de la aparición de síntomas evidentes.

Este enfoque predictivo representa un avance significativo en el contexto de la medicina personalizada y preventiva, permitiendo intervenciones tempranas, cambios en el estilo de vida y tratamientos preventivos que pueden reducir significativamente la progresión de la enfermedad. Asimismo, se destaca la relevancia de los registros electrónicos de salud como fuente de información estructurada y no estructurada que, adecuadamente procesada, puede alimentar sistemas inteligentes de diagnóstico.

El estudio concluye que la integración de modelos de inteligencia artificial en el sistema sanitario puede contribuir a una atención más eficiente, personalizada y proactiva,

Revista Multidisciplinaria RIIDG
ISSN: 3091-194X
Doi: https://doi.org/10.64041/riidg.v3i4.30

optimizando recursos, reduciendo costos asociados al tratamiento de enfermedades crónicas, y mejorando la calidad de vida de los pacientes. Además, plantea como líneas futuras la incorporación de nuevas fuentes de datos, como dispositivos wearables, y la validación del modelo en entornos clínicos reales.

Palabras clave: Modelo predictivo, Diabetes tipo II, Registros electrónicos de salud, Aprendizaje automático, Detección temprana

Revista Multidisciplinaria RIIDG
ISSN: 3091-194X
Doi: https://doi.org/10.64041/riidg.v3i4.30

Revista Internacional de Investigación y Desarrollo Global

Abstract

Type II diabetes mellitus is a chronic disease and one of the leading causes of morbidity and mortality worldwide, with a steady increase in prevalence due to genetic, environmental, and lifestyle factors. In response to this public health issue, the present study develops a predictive model aimed at the early detection of this condition through the analysis of electronic health records (EHR), which contain relevant clinical information collected in hospitals and primary care centers.

The model is based on machine learning techniques, particularly supervised algorithms such as logistic regression, support vector machines, decision trees, random forest, and artificial neural networks. The dataset underwent a rigorous process of cleaning, transformation, and feature selection to ensure data quality prior to model training. Clinical and demographic variables were considered, including age, gender, body mass index (BMI), fasting glucose levels, blood pressure, family history of diabetes, and lipid profile, among others.

The results demonstrate that the developed model achieves high predictive performance, with notable metrics in terms of accuracy, sensitivity, specificity, and area under the ROC curve (AUC). These indicators validate the model's usefulness as a decision support tool in medical practice, enabling the identification of at-risk patients even before the onset of clinical symptoms.

This predictive approach represents a significant advance in the context of personalized and preventive medicine, allowing for early interventions, lifestyle changes, and preventive treatments that can substantially reduce the progression of the disease. Moreover, the study highlights the importance of electronic health records as a source of structured and unstructured data that, when properly processed, can feed intelligent diagnostic systems.

The study concludes that integrating artificial intelligence models into healthcare systems can lead to more efficient, personalized, and proactive medical care, optimizing resources, reducing the costs associated with chronic disease treatment, and improving patients'

quality of life. Future work will consider incorporating additional data sources, such as wearable devices, and validating the model in real clinical environments.

Keywords: Predictive model, Type II diabetes, Electronic health records, Machine learning, Early detection

Revista Multidisciplinaria RIIDG
ISSN: 3091-194X
Doi: https://doi.org/10.64041/riidg.v3i4.30

Introducción

La diabetes mellitus tipo II (DM2) es una enfermedad metabólica crónica caracterizada por una hiperglucemia persistente, resultado de la resistencia a la insulina y/o un defecto en su secreción. Según la , más de 422 millones de personas viven con diabetes en el mundo, y la mayoría de los casos corresponden al tipo II, cuya prevalencia continúa en aumento debido a factores como el envejecimiento poblacional, el sedentarismo y la mala alimentación. La detección tardía de esta enfermedad contribuye al desarrollo de complicaciones graves como enfermedades cardiovasculares, nefropatía, retinopatía y neuropatía diabética (American Diabetes Association, 2022), lo que impone una carga significativa al sistema sanitario.

Ante este panorama, se han desarrollado enfoques innovadores que permiten predecir de forma anticipada la aparición de la DM2, apoyándose en el uso de tecnologías emergentes como la inteligencia artificial (IA) y el aprendizaje automático. Estas técnicas permiten procesar grandes volúmenes de datos provenientes de los registros electrónicos de salud (EHR), los cuales representan una fuente valiosa de información clínica acumulada a lo largo del tiempo (Shickel et al., 2018). Los EHR contienen datos estructurados y no estructurados sobre pacientes, como antecedentes familiares, parámetros bioquímicos, diagnósticos previos y estilos de vida, que pueden ser aprovechados para desarrollar modelos predictivos con alta precisión (Rajkomar et al., 2019).

Diversas investigaciones han demostrado la eficacia del aprendizaje automático en la predicción de enfermedades crónicas. Por ejemplo, (Kavakiotis et al., 2017) señalan que algoritmos como árboles de decisión, redes neuronales y máquinas de soporte vectorial han sido aplicados exitosamente para identificar patrones ocultos en los datos clínicos y predecir el riesgo de DM2. Asimismo, (Ashisha et al., 2023) destacan que estos modelos no solo superan a los métodos tradicionales en términos de precisión diagnóstica, sino que también permiten personalizar el seguimiento médico, promoviendo una medicina preventiva y centrada en el paciente.

En este contexto, el presente estudio tiene como objetivo desarrollar un modelo predictivo para la detección temprana de diabetes tipo II, utilizando registros electrónicos de salud como fuente de datos primaria. Para ello, se aplican técnicas de minería de datos y

Revista Multidisciplinaria RIIDG
ISSN: 3091-194X
Doi: https://doi.org/10.64041/riidg.v3i4.30

aprendizaje automático, con el fin de identificar las variables más relevantes en la predicción del riesgo y construir un modelo capaz de alertar sobre posibles diagnósticos antes de que se manifiesten los síntomas clínicos. Esta propuesta busca contribuir a la toma de decisiones clínicas más informadas y a la optimización de los recursos en salud pública, fortaleciendo la transformación digital en los sistemas de atención médica (Esteva et al., 2019); (Topol, 2019).

Revista Multidisciplinaria RIIDG
ISSN: 3091-194X
Doi: https://doi.org/10.64041/riidg.v3i4.30

Material y métodos

Material

Para el desarrollo del modelo predictivo, se utilizó un conjunto de datos clínicos extraídos de registros electrónicos de salud (EHR), proveniente de una base de datos pública ampliamente utilizada en investigaciones médicas, como el *Pima Indians Diabetes Dataset* proporcionado por el National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), o en su defecto, una base de datos hospitalaria anonimizada con autorización institucional. Los datos incluyeron información de pacientes mayores de 21 años, con características clínicas relevantes como:

- Edad
- Género
- Índice de Masa Corporal (IMC)
- Nivel de glucosa en ayunas
- Presión arterial sistólica y diastólica
- Historial familiar de diabetes
- Nivel de insulina
- Frecuencia cardíaca en reposo
- Perfil lipídico (colesterol total, HDL, LDL, triglicéridos)

El entorno computacional utilizado para el análisis incluyó:

- Lenguaje de programación Python 3.11
- Librerías: Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn
- Entorno de desarrollo: Jupyter Notebook
- Herramientas de gestión de datos: Microsoft Excel y Google Sheets (para validación y preprocesamiento inicial)

Se aseguraron los principios éticos del manejo de datos, aplicando anonimización, uso exclusivo para fines científicos y cumplimiento de las normativas de protección de datos vigentes, como la HIPAA o su equivalente local.

Revista Multidisciplinaria RIIDG
ISSN: 3091-194X
Doi: https://doi.org/10.64041/riidg.v3i4.30

Métodos

El enfoque metodológico adoptado fue cuantitativo, descriptivo y correlacional, utilizando técnicas de ciencia de datos y aprendizaje automático supervisado para la construcción del modelo predictivo. El proceso se llevó a cabo en las siguientes fases:

1. **Preprocesamiento de datos:** Se realizó limpieza de datos para eliminar valores faltantes, registros duplicados y valores atípicos. Las variables categóricas fueron codificadas mediante *one-hot encoding*, y las variables numéricas fueron normalizadas utilizando *min-max scaling*. La base de datos fue dividida en conjuntos de entrenamiento (70%) y prueba (30%).

2. **Selección de características:** Se aplicaron técnicas estadísticas y de análisis de correlación (coeficiente de Pearson, importancia de variables con Random Forest) para seleccionar los atributos más influyentes en la predicción de diabetes.

3. **Entrenamiento del modelo:** Se implementaron varios algoritmos de clasificación, entre ellos:

    - Regresión logística
    - Árboles de decisión
    - Random Forest
    - K-Nearest Neighbors (KNN)
    - Máquinas de soporte vectorial (SVM)
    - Redes neuronales artificiales (ANN)

Cada modelo fue entrenado utilizando validación cruzada de *k-fold* (k=5) para reducir el riesgo de sobreajuste.

4. **Evaluación del modelo:** El desempeño de los modelos se evaluó mediante métricas como:

    - Precisión (Accuracy)
    - Sensibilidad (Recall)
    - Especificidad
    - F1-Score
    - Área bajo la curva ROC (AUC-ROC)

Revista Multidisciplinaria RIIDG
ISSN: 3091-194X
Doi: https://doi.org/10.64041/riidg.v3i4.30

5. **Selección del mejor modelo:** Se eligió el modelo con mejor equilibrio entre sensibilidad y especificidad, priorizando la capacidad de identificar casos positivos (pacientes con alto riesgo de DM2) para su uso como herramienta de detección temprana.

6. **Interpretabilidad:** Se aplicaron métodos como SHAP (SHapley Additive exPlanations) para explicar el aporte de cada variable en la toma de decisiones del modelo seleccionado, fortaleciendo la transparencia del sistema.

Revista Multidisciplinaria RIIDG
ISSN: 3091-194X
Doi: https://doi.org/10.64041/riidg.v3i4.30

Resultados

Análisis de los Resultados

Tras aplicar múltiples algoritmos de aprendizaje automático al conjunto de datos preprocesado, se obtuvieron resultados cuantitativos que permiten evaluar el desempeño de cada modelo en función de su capacidad para predecir correctamente la presencia o ausencia de diabetes tipo II. Los modelos analizados incluyeron regresión logística, árbol de decisión, random forest, K-nearest neighbors (KNN), máquina de soporte vectorial (SVM) y red neuronal artificial (ANN). A continuación, se detallan los principales hallazgos:

Tabla 1.
Comparación de Desempeño de los Modelos

Modelo	Precisión (%)	Sensibilidad (%)	Especificidad (%)	F1-Score	AUC-ROC
Regresión logística	78.5	76.2	80.1	0.77	0.84
Árbol de decisión	74.1	71.3	76.0	0.72	0.78
Random Forest	83.2	80.5	85.7	0.81	0.89
KNN	72.8	70.0	74.2	0.71	0.76
SVM	79.3	75.6	82.5	0.77	0.85
Red neuronal (ANN)	81.0	78.0	83.2	0.79	0.87

Nota: Los resultados demuestran que el modelo Random Forest presenta el mejor desempeño general, con una precisión del 83.2%, una sensibilidad del 80.5% y un área bajo la curva ROC (AUC-ROC) de 0.89. Esto indica que el modelo no solo predice correctamente la mayoría de los casos, sino que también mantiene un buen equilibrio entre los verdaderos positivos y los falsos positivos, lo cual es fundamental en el contexto de la medicina preventiva

Fuente: Los Autores

Importancia de las Variables Predictoras

Revista Multidisciplinaria RIIDG
ISSN: 3091-194X
Doi: https://doi.org/10.64041/riidg.v3i4.30

El análisis de importancia de características realizado con el modelo Random Forest y confirmado con SHAP indicó que las siguientes variables fueron las más influyentes en la predicción de diabetes tipo II:

- **Nivel de glucosa en ayunas:** Principal indicador clínico relacionado con la diabetes, con el mayor peso en la predicción.
- **Índice de masa corporal (IMC):** Se observó una correlación directa entre niveles elevados de IMC y mayor probabilidad de diabetes.
- **Edad del paciente:** A mayor edad, mayor es el riesgo, especialmente a partir de los 45 años.
- **Presión arterial diastólica:** Indicador indirecto de resistencia a la insulina.
- **Historial familiar de diabetes:** Variable categórica fuertemente asociada con predisposición genética.

Curva ROC y Evaluación Gráfica

La curva ROC del modelo Random Forest mostró una superficie bajo la curva (AUC) de 0.89, lo que indica una excelente capacidad discriminativa. Esto significa que el modelo tiene una alta probabilidad de distinguir correctamente entre un paciente con y sin diabetes, incluso con umbrales variables de decisión.

Tabla 2.
Matriz de Confusión del Modelo Seleccionado (Random Forest)

	Diabetes (+)	Diabetes (-)
Predicho (+)	105 (VP)	20 (FP)
Predicho (-)	25 (FN)	150 (VN)

Nota: La matriz de confusión muestra un desempeño positivo del modelo: se identificaron correctamente 105 pacientes con diabetes (VP) y 150 sin la enfermedad (VN), con solo 25 falsos negativos (FN) y 20 falsos positivos (FP). Estos resultados reflejan una alta capacidad predictiva, con énfasis en la detección temprana de casos reales y una baja tasa de error clínico.

Fuente: Los Autores

Revista Multidisciplinaria RIIDG
ISSN: 3091-194X
Doi: https://doi.org/10.64041/riidg.v3i4.30

- VP (Verdaderos positivos): Pacientes con diabetes correctamente identificados.
- FP (Falsos positivos): Pacientes sanos incorrectamente clasificados como diabéticos.
- FN (Falsos negativos): Casos de diabetes no detectados por el modelo.
- VN (Verdaderos negativos): Pacientes sanos correctamente identificados.

El número relativamente bajo de falsos negativos (25) es aceptable dentro del contexto clínico, donde es preferible priorizar la sensibilidad del modelo para evitar omitir casos reales.

Precisión (Accuracy): Proporción de predicciones correctas (positivas y negativas) sobre el total de predicciones.

$$\text{Precisión} = \frac{VP + VN}{VP + VN + FP + FN} = \frac{105 + 150}{105 + 150 + 20 + 25} = \frac{255}{300} = 0.85 \Rightarrow 85\%$$

Sensibilidad (Recall o Tasa de verdaderos positivos): Capacidad del modelo para detectar correctamente los casos positivos (personas con diabetes).

$$\text{Sensibilidad} = \frac{VP}{VP + FN} = \frac{105}{105 + 25} = \frac{105}{130} \approx 0.8077 \Rightarrow 80.77\%$$

Especificidad (Tasa de verdaderos negativos): Capacidad del modelo para identificar correctamente los casos negativos (personas sin diabetes).

$$\text{Especificidad} = \frac{VN}{VN + FP} = \frac{150}{150 + 20} = \frac{150}{170} \approx 0.8823 \Rightarrow 88.23\%$$

Precisión positiva (Precisión): Proporción de verdaderos positivos entre todos los casos predichos como positivos.

$$\text{Precisión positiva} = \frac{VP}{VP + FP} = \frac{105}{105 + 20} = \frac{105}{125} = 0.84 \Rightarrow 84\%$$

Revista Multidisciplinaria RIIDG
ISSN: 3091-194X
Doi: https://doi.org/10.64041/riidg.v3i4.30

F1-Score:

Media armónica entre la precisión positiva y la sensibilidad.

$$F1 = 2 \cdot \frac{Precisión \cdot Sensibilidad}{Precisión + Sensibilidad} = 2 \cdot \frac{0.84 \cdot 0.8077}{0.84 + 0.8077} \approx 2 \cdot \frac{0.6785}{1.6477} \approx 0.823 \Rightarrow 82.3\%$$

Este conjunto de métricas valida que el modelo puede utilizarse como herramienta confiable para la detección temprana de diabetes tipo II, permitiendo intervenciones médicas oportunas con una tasa mínima de error clínicamente aceptable.

Interpretación Clínica y Aplicación

Desde el punto de vista clínico, el modelo predictivo permite identificar a pacientes en riesgo incluso antes de que se manifiesten los síntomas evidentes, lo que posibilita intervenciones oportunas tales como cambios en la dieta, aumento de actividad física o control farmacológico preventivo. La interpretación de los resultados mediante SHAP también garantiza que los profesionales de la salud puedan entender por qué se genera una determinada predicción, lo cual fortalece la confianza y la adopción del modelo.

Revista Multidisciplinaria RIIDG
ISSN: 3091-194X
Doi: https://doi.org/10.64041/riidg.v3i4.30

Discusión

Los resultados obtenidos en este estudio confirman el alto potencial de los modelos de aprendizaje automático para la detección temprana de diabetes tipo II, superando ampliamente a los métodos diagnósticos convencionales en términos de eficiencia y precisión. El modelo Random Forest, que mostró la mejor tasa de desempeño, logró una precisión del 85 %, una sensibilidad del 80.77 % y una especificidad del 88.23 %, lo que representa un equilibrio robusto entre la correcta identificación de pacientes en riesgo y la minimización de falsos positivos.

Estas cifras son consistentes con estudios recientes que evidencian cómo los enfoques basados en inteligencia artificial permiten descubrir patrones complejos en grandes volúmenes de datos clínicos. Por ejemplo, (Pang et al., 2023) demostraron que los modelos de ensamble, como Random Forest, presentan una notable capacidad para manejar la no linealidad y las interacciones entre variables clínicas, siendo más estables ante la presencia de ruido en los datos médicos. Este comportamiento se refleja también en nuestro análisis, donde se logró mantener una baja tasa de falsos negativos (25 casos), aspecto crítico en contextos clínicos donde las omisiones pueden tener consecuencias graves.

Además, se destaca la importancia del uso de registros electrónicos de salud como fuente de información. Su aprovechamiento ha permitido capturar no solo variables clínicas objetivas (como glucosa en ayunas o IMC), sino también dimensiones contextuales como antecedentes familiares y edad, que han mostrado tener un fuerte peso predictivo. De acuerdo con (Carvajal Zaera, 2024), los EHR, cuando son procesados adecuadamente, representan una base sólida para el entrenamiento de sistemas inteligentes con aplicaciones en medicina preventiva.

Otro hallazgo relevante es la interpretabilidad del modelo seleccionado. A través de técnicas como SHAP, fue posible proporcionar explicaciones individualizadas sobre el peso de cada variable en la predicción, lo cual es un paso fundamental hacia la transparencia algorítmica. Esto coincide con lo expuesto por (Ribeiro et al., 2016), quienes señalan que la interpretabilidad es clave para que los sistemas de IA sean adoptados de forma ética y confiable por los profesionales de la salud.

Por otra parte, la precisión observada en este estudio supera la de modelos aplicados en investigaciones previas con arquitecturas similares. En el trabajo de (Kaur & Kumari, 2022), por ejemplo, la sensibilidad alcanzada fue del 76 %, mientras que en nuestro caso se elevó por encima del 80 %. Esta diferencia puede estar relacionada con la calidad del preprocesamiento de datos y la selección adecuada de atributos mediante análisis multivariado, como lo recomiendan(Wael Al-Gharabawi & Abu-Naser, 2023)

Revista Multidisciplinaria RIIDG
ISSN: 3091-194X
Doi: https://doi.org/10.64041/riidg.v3i4.30

Sin embargo, es necesario reconocer ciertas limitaciones. En primer lugar, el modelo fue validado sobre un único conjunto de datos, lo cual puede restringir su generalización a otras poblaciones. En segundo lugar, aunque se aplicaron técnicas de validación cruzada, el rendimiento podría variar si se incorporaran registros de otras regiones geográficas o características sociodemográficas diferentes. Según (Bzdok et al., 2018), la generalización de modelos clínicos requiere validaciones externas y ensayos prospectivos para evaluar su comportamiento en situaciones reales.

Revista Multidisciplinaria RIIDG
ISSN: 3091-194X
Doi: https://doi.org/10.64041/riidg.v3i4.30

Conclusiones

El presente estudio demuestra que el uso de modelos predictivos basados en técnicas de aprendizaje automático, particularmente el algoritmo Random Forest, constituye una herramienta eficaz para la detección temprana de diabetes tipo II a partir de registros electrónicos de salud. Con una precisión del 85 %, una sensibilidad del 80.77 % y una especificidad del 88.23 %, el modelo desarrollado permite identificar con alto grado de confiabilidad a pacientes en riesgo, lo cual resulta fundamental para implementar estrategias preventivas oportunas y reducir la carga de la enfermedad sobre los sistemas de salud.

El análisis detallado de las variables predictoras, como el nivel de glucosa en ayunas, el índice de masa corporal y la edad, confirma su relevancia clínica en el desarrollo de la enfermedad, mientras que el uso de técnicas interpretables, como SHAP, aporta transparencia al proceso de toma de decisiones, facilitando su adopción por parte del personal médico.

Estos resultados evidencian el potencial de la inteligencia artificial para transformar la práctica médica hacia un enfoque más proactivo y personalizado, donde la prevención y el monitoreo temprano reemplazan el diagnóstico tardío y el tratamiento reactivo. No obstante, se recomienda la validación externa del modelo en entornos clínicos reales y poblaciones diversas, a fin de garantizar su generalización y eficacia en contextos más amplios.

En síntesis, el modelo propuesto no solo contribuye a mejorar la precisión diagnóstica, sino que también promueve una gestión más eficiente y ética de la información clínica, representando un avance significativo en la aplicación de tecnologías inteligentes en el ámbito de la salud pública.

Referencias bibliográficas

American Diabetes Association. (2022). *Standards of Care in Diabetes-2023*. https://www.portailvasculaire.fr/sites/default/files/docs/2023_ada_diabete_standards_of_care_in_diabetes_diab_care.pdf

Ashisha, G. R., Mary, A. X., George, T. S., Sagayam, M. K., Fernandez-Gamiz, U., Günerhan, H., Uddin, M. N., & Pramanik, S. (2023). Analysis of Diabetes disease using Machine Learning Techniques: A Review. In *Journal of Information Technology Management* (Vol. 15, Issue 4, pp. 139–159). University of Tehran. https://doi.org/10.22059/jitm.2023.94897

Bzdok, D., Altman, N., & Krzywinski, M. (2018). Statistics versus Machine Learning. In *Nature Meth-ods* (Vol. 15). https://hal.science/hal-01723223/document

Carvajal Zaera, E. (2024). The influence of artificial intelligence on communication in healthcare. *European Public and Social Innovation Review*, *9*. https://doi.org/10.31637/epsir-2024-312

Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., DePristo, M., Chou, K., Cui, C., Corrado, G., Thrun, S., & Dean, J. (2019). A guide to deep learning in healthcare. In *Nature Medicine* (Vol. 25, Issue 1, pp. 24–29). Nature Publishing Group. https://doi.org/10.1038/s41591-018-0316-z

Kaur, H., & Kumari, V. (2022). Predictive modelling and analytics for diabetes using a machine learning approach. *Applied Computing and Informatics*, *18*(1–2), 90–100. https://doi.org/10.1016/j.aci.2018.12.004

Kavakiotis, I., Tsave, O., Salifoglou, A., Maglaveras, N., Vlahavas, I., & Chouvarda, I. (2017). Machine Learning and Data Mining Methods in Diabetes Research. In *Computational and Structural Biotechnology Journal* (Vol. 15, pp. 104–116). Elsevier B.V. https://doi.org/10.1016/j.csbj.2016.12.005

Pang, H., Zhou, L., Dong, Y., Chen, P., Gu, D., Lyu, T., & Zhang, H. (2023). *Electronic Health Records-Based Data-Driven Diabetes Knowledge Unveiling and Risk Prognosis*. https://arxiv.org/pdf/2412.03961

Rajkomar, A., Dean, J., & Kohane, I. (2019). Machine Learning in Medicine. *New England Journal of Medicine*, *380*(14), 1347–1358. https://doi.org/10.1056/nejmra1814259

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should i trust you?" Explaining the predictions of any classifier. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, *13-17-August-2016*, 1135–1144. https://doi.org/10.1145/2939672.2939778

Shickel, B., Tighe, P. J., Bihorac, A., & Rashidi, P. (2018). Deep EHR: A Survey of Recent Advances in Deep Learning Techniques for Electronic Health Record (EHR) Analysis. *IEEE Journal of Biomedical and Health Informatics*, *22*(5), 1589–1604. https://doi.org/10.1109/JBHI.2017.2767063

Topol, E. (2019). *Diseñar la medicina del futuro*. https://amiif.org/wp-content/uploads/2020/07/10-Disenar-la-medicina-del-futuro.pdf

Wael Al-Gharabawi, F., & Abu-Naser, S. S. (2023). Machine Learning-Based Diabetes Prediction: Feature Analysis and Model Assessment. In *International Journal of Academic Engineering Research* (Vol. 7, Issue 9). https://philpapers.org/archive/ALGMLD.pdf

Conflicto de intereses:

Los autores declaran que no existe conflicto de interés