CAPÍTULO 13

ESTRATEGIA TECNOLÓGICA INTELIGENTE DE APOYO A LA PREVENCIÓN DE ENFERMEDADES CARDIACAS

Data de aceite: 02/08/2023

Silvia Soledad Moreno Gutiérrez
Profesor investigador en la Universidad Autónoma del Estado de Hidalgo. México.

Héctor Daniel Molina Ruíz
Profesor investigador en la Universidad Autónoma del Estado de Hidalgo. México.

El aprendizaje automático mejor conocido como machine learning, ha ofrecido amplio apoyo y beneficios a los sectores de la sociedad. En el área de la salud su enfoque ha sido orientado hacia los grupos vulnerables por su necesidad de atención y de solucionar problemáticas potenciadas por su situación económica generalmente desfavorable que les impide el acceso a la atención médica. Al respecto, el presente trabajo tiene como propósito construir una estrategia de prevención de enfermedades del corazón, considerando que están ubicadas como la primera causa de muerte en el mundo y también en México. Por lo anterior, se desarrolló una estrategia tecnológica basada en un modelo de predicción de enfermedades en el corazón a través de aprendizaje automático y la técnica de Inteligencia Artificial (AI) conocida como Redes Neuronales Artificiales (RNA). Considerando el impacto favorable que el aprendizaje automático ha significa para el área de la salud por sus diversas propuestas desarrolladas para pronóstico de enfermedades, y con base en el problema derivado de los problemas cardiacos cada vez más frecuentes en el mundo y con alto índice de mortalidad, se decidió ofrecer apoyo a través del desarrollo tecnológico.

Para ello, una de las metodologías de ciencia de datos de mayor aplicación es la Proceso Estándar Entre Industrias para la Minería de Datos (CRISP-DM por sus siglas en inglés) orientadas a la construcción de modelos a través del aprendizaje automático basado en datos. Se construyeron tres modelos de clasificación, el de mejor rendimiento fue el perceptron multicapa con 93.23% de precisión, 91.65% de exactitud, 91.85% de especificidad y 91.49% de sensibilidad, por lo que se consideró adecuado para

apoyar el diagnóstico oportuno del paciente. El trabajo que se expone muestra el proceso de desarrollo de la estrategia tecnológica.

INTRODUCCIÓN

Las enfermedades del corazón representan una de las principales causas de muerte para los habitantes en el planeta, en países de Norte América estas patologías se presentan con frecuencia y el 47% tiene asociado algún factor de riesgo como presión arterial alta o colesterol alto. Estas enfermedades son crónicas, es decir, permanecen por largos periodo de tiempo, su causa no es precisa y en general no tienen cura, según la Organización Mundial de la Salud (OMS) constituyen un problema grave para la sociedad en cada región del mundo y son responsables de 17.9 millones de muertes por año (OMS, 2022).

Con base en esta problemática y considerando el alto potencial de la ciencia de datos y del aprendizaje automático en el ámbito de la salud, de manera permanente se realiza revisión del estado del arte en cuanto a las propuestas publicadas y disponibles, cuyo propósito es contribuir a la prevención de estas enfermedades a través de la construcción de estrategias tecnológicas de fácil uso y accesibilidad.

La literatura expone amplio número de propuestas e investigaciones con el propósito que se aborda.

Llodrá (2018) Desarrolló un algoritmo para la clasificación de arritmias cardíacas utilizando el modelo de Redes Neuronales Convolucionales (RNC) con una precisión de 67.09%. Cucas et al (2021) Desarrollaron un algoritmo para la clasificación de arritmias cardiacas utilizando la técnica de señales provenientes de la realización de Electro Encefalogramas (EGG9, aplicando algoritmo de la máquina de soporte vectorial (MSV), K-Vecinos más cercanos (KNN), Perceptrón Multicapa (MLP) y Árbol de decisión (AD).

Pérez Soria (2019) desarrolló algoritmo para identificar factores de riesgos de insuficiencia cardiaca con Random Forest (RF) y alcanzó 35%, AdaBoost con 33% y MLP con 37%. Confident (2018) aplicó algoritmo para diagnosticar enfermedades cardiacas ejecutando mediante MSV con precisión del 64.36%. Choque Forra y Mamani Mamani (2020) Desarrollaron un predictor de supervivencia ante la insuficiencia cardiaca utilizando Regresión Lineal (RL) de 87% de precisión, K-NN con 67%, AD con 90%, MSV con 72% y RF con 90%, siendo este último la mejor alternativa.

Ayala Poma y Huaman Ollero (2020) Desarrollaron un algoritmo de predicción de complicaciones cardiacas utilizando minería de datos y señales electro cardiográficas, lograron precisión de 90%, base de datos clínicos con 85% y con aprendizaje automático los dispositivos open source con 70% y Wearables con 50%, mostraba más técnicas las cuales no tienen una precisión establecida, estas técnicas utilizaron un entrenamiento supervisado, semi-supervisado y no supervisado.

Javier (2019) aplicó algoritmo inteligente para clasificación del audio cardiaco como

la Regresión logística (LR) con 76% de precisión, Análisis Discriminante Lineal (LDA) con 72%, K-NN con 75%, AD con 74% y redes bayesianas con 63%. Zapana (2021) Desarrolló un algoritmo para la determinación de valvulopatías cardiacas a través del análisis de sonidos del corazón con modelos de CNN con una precisión de 89.7% en diagnóstico de pacientes sanos y 78.9% en pacientes con valvulopatías. Gallego Valcárcel y Lucas Monsalve (2021) Desarrollaron un algoritmo para predecir el riesgo de fatalidad por insuficiencia cardiaca mediante MSV con 86.51% de precisión, MLP con 87.38% y RF con 83.67%.

Mohan et al., (2019) Desarrollaron un predictor para enfermedades cardiacas usando técnicas híbridas de aprendizaje automático siendo como mejor acertada los Árboles Potenciado por Gradiente con un 94.1% de precisión, Bayer ingenuo con 90.5% y RL con 89,9%. Bharti et al (2021) Desarrollaron predicción de enfermedades cardiacas mediante una combinación de aprendizaje automático y aprendizaje profundo utilizando RF con precisión de 80.3%, RL con 83,31%, K-NN con 84.86%, MSV con 83.29%, AD con 83.33%, XGBoost con 71.4%. Dangare y Apte en el artículo de Aljanabi et al., (2018) Desarrollaron un algoritmo para la predicción de enfermedades del corazón utilizando RNA con precisión de 80%, sensibilidad 85% y especificidad de 70%.

Dada la importancia de brindar estrategias que contribuyan al diagnóstico y prevención de estas enfermedades, así como los resultados de la revisión efectuada que expresan unos resultados que aun requieren estudios e investigación exhaustiva, el presente trabajo consiste en una propuesta basada en RNA para predecirlas, la cual constituirá un apoyo para la prevención de este problema de salud. En la presente propuesta se aplicaron otros algoritmos de aprendizaje tales como RNA, RL y RF para obtener un criterio más amplio basado en el análisis comparativo.

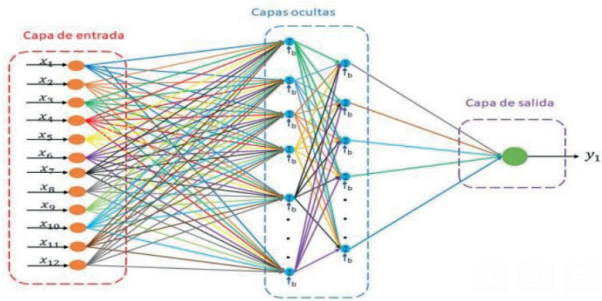

Figura1. RNA

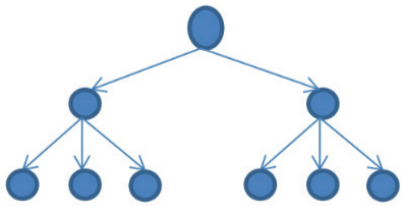

Figura2. RF

Desarrollo de un predictor de enfermedades cardiacas

Se inició el trabajo a partir de la descarga del banco de datos que se extrajo de la página kaggle.com el cual integra los indicadores clave de personas con enfermedades cardíacas, se contó con un total de 31,979 registros y 14 variables que se exponen en la tabla 1.

Se aplicó la Metodología de Proceso Estándar de la Industria Cruzada para la Minería de Datos (CRISP-DM, por sus siglas en inglés), adecuada para explotación de datos mediante técnicas de aprendizaje automático. Las técnicas aplicadas fueron RNA, RF y RL por su habilidad y potencial de predicción.

Metodología CRISP-DM

Esta metodología consta de 6 pasos que se desarrollaron y se exponen a continuación:

1. Entendimiento del negocio
2. Entendimiento de los datos
3. Preparación de los datos
4. Elección del modelo
5. Validación del modelo
6. Despliegue (Arias et al., 2021).

En la fase 1 analizó la situación actual relacionada con las enfermedades cardiacas y problema que representa a nivel mundial, de igual forma, se analizó el hecho de que los modelos de predicción hoy en día apoyan de forma importante las áreas de la salud, de igual forma, dado que las redes neuronales artificiales ofrecen alto potencial en tareas de este tipo, se aplicaron a la solución que se busca.

En la fase 2, se analizaron los datos reunidos en el banco de datos el cual consta de las variables que se muestran en la tabla1.

Variable	Tipo
¿Enfermo?	Categórica
IMC	Numérica
Estado de fumador	Categórica
Alcohólico	Categórica
Derrame cerebral	Categórica
Salud física (Rango del 1 al 100)	Numérica
Salud mental (Cuantos días, llega hasta 30)	Numérica
Camina con problema	Categórica
Diabético?	Categórica
Actividad Física	Categórica
Salud General	Categórica
Horas que duerme	Numérica
Asma	Categórica
Enfermedad del Riñón	Categórica

Tabla 1. Variables

El banco de datos contó con registros completos y sin datos nulos. El análisis multivariado mediante la matriz de correlación permitió identificar la alta correlación entre las variables, enfatizando en salud general y dificultades para caminar, las variables seleccionadas fueron 14.

En la fase 3, se eliminaron variables irrelevantes: raza, sexo, cáncer de piel. La variable objetivo es ¿enfermo? es categórica por lo que se desarrolló un modelo de clasificación. Posteriormente se generaron “dummies” en algunas variables con más de dos categorías, por ejemplo la variable Salud general y salud mental. Posteriormente se realizó la normalizando los datos con *MinMax scaler*.

En la fase 4 se entrenaron algunos modelos con el propósito de elegir el mejor. Se dividió el banco de datos en dos bloques, el de entrenamiento (80%) y el de prueba (20%). Se entrenaron 3 modelos a partir de los algoritmos de aprendizaje antes mencionados.

Se aplicó algoritmo de red neuronal RNA con MLP, RF Y RL, logrando los mejores resultados el segundo modelo.

En la fase de validación se aplicó matriz de confusión con las siguientes ecuaciones (Düntsch y Gediga, 2019). Para conseguir los resultados se calcularon:

Precisión	VP/(VP+FP)
Exactitud	VP+VN/(VP+FP+FN+VN)
Sensibilidad	VP/(VP+FN)
Especificidad	VN/(VN+FP)

Comportamiento del modelo ante el problema

Los modelos aplicados a la solución del problema así como sus métricas de rendimiento se muestran a continuación como resultado en la tabla 2, tabla3 y tabla 4.

Verdaderos Positivos (VP) 3224	Falsos Positivos (FP) 234
Falsos Negativos (FN) 300	Verdaderos Negativos (VN) 2638

Tabla 2. Resultados MLP

Verdaderos Positivos (VP) 3454	Falsos Positivos (FP) 256
Falsos Negativos (FN) 384	Verdaderos Negativos (VN) 2302

Tabla 3. Resultados RL

Verdaderos Positivos (VP) 3328	Falsos Positivos (FP) 467
Falsos Negativos (FN) 382	Verdaderos Negativos (VN) 2219

Tabla 4. Resultados RF

La tabla 5 muestra el resumen de resultados en cada caso, a partir de la cual es posible observar el rendimiento de cada uno, considerando un total de 25,583 registros para entrenamiento y 6396 para validación.

Modelo	Métrica de validación	Precisión	Exactitud	Sensibilidad	Especificidad
MLP	Matriz de confusión	93.23%	91.65%	91.49%	91.85%
RL		93.10%	89.99%	89.99%	89.99%
RF		87.69%	86.73%	89.70%	82.61%

Tabla 5. Resultados

COMENTARIOS FINALES

El mejor rendimiento se obtuvo a través del MLP luego de entrenar el modelo de RL y RF. El mejor modelo logró representar de forma adecuada los patrones de entrada durante el entrenamiento para luego predecir con alta precisión la enfermedad en los pacientes. Las RNA son alternativas tecnológicas de alto potencial para la predicción en el área de salud.

El aprendizaje automático ha brindado apoyo y adquirido protagonismo en los últimos años debido a sus posibilidades de representar problemas de comportamiento no lineal, y con base en esto ha logrado incrementar la calidad de vida de las persona.

Las técnicas de machine learning han sido ampliamente aplicadas en el diagnóstico de enfermedades cardiacas, no obstante, han alcanzado resultados de precisión que aun requieren estudios de mayor profundidad, sin embargo, el modelo que se propone alcanzó un rendimiento superior a los identificados en la literatura.

REFERENCIAS

Aljanabi, M., Qutqut, M. H., & Hijjawi, M. Machine learning classification techniques for heart disease prediction: a review. International Journal of Engineering & Technology, 7(4), 5373-5379. 2018.

Ayala Poma, M. E., & Huaman Ollero, J. A. Técnicas y Herramientas para la predicción de complicaciones cardiacas, utilizando wearables inteligentes: una revisión sistemática de la literatura.2020.

Arias, E. B. N., Nuñez, B. M. G., Fernández, L. N., & Pupo, J. M. R. CRISP-DM y K-means neutrosofía en el análisis de factores de riesgo de pérdida de audición en niños. Revista Asociación Latinoamericana de Ciencias Neutrosóficas. ISSN 2574-1101, 16, 73-81. (2021).

Bharti, R., Khamparia, A., Shabaz, M., Dhiman, G., Pande, S., & Singh, P. . Prediction of heart disease using a combination of machine learning and deep learning. Computational intelligence and neuroscience, 2021.

Choque Forra, D. P., & Mamani Mamani, J. L. Predicción de supervivencia ante la insuficiencia cardíaca. Revista de Investigación Estudiantil Iluminate, 12, 77. 2020.

CONFIDENT, L. Estrategia “muestra del menos confiable” para el diagnóstico de enfermedades cardiacas. (2018).

Cucas, H. A. A., Piscal, E. A. M., Torres, D. M., & Chamorro, A. X. O. Diseño de un sistema de procesamiento y caracterización de potenciales ECG para la clasificación de arritmias cardiacas, mediante el uso de técnicas de aprendizaje automático supervisadas. Boletín Informativo CEI, 8(2), 204-210. 2021.

Düntsch, I., & Gediga, G. (2019). Confusion matrices and rough set data analysis. In Journal of Physics: Conference Series (Vol. 1229, No. 1, p. 012055). IOP Publishing.

Gallego Valcárcel, D. A. & Lucas Monsalve, D. F. Modelos de aprendizaje automático para la predicción del riesgo de fatalidad por insuficiencia cardiaca con datos clínicos. Repositorio. UAN. Recuperado 7 de septiembre de 2022, de http://repositorio.uan.edu.co/bitstream/123456789/4803/5/2021DavidAlejandroGallegoMonografi%CC%81a.pdf

Hoyos, M. L., Vivas, M. C. B., & López, J. M. L. Detección Automática De Soplos Cardiacos A Partir De La Señal De Fonocardiografía. Encuentro Internacional de Educación en Ingeniería. 2019.

Javier, I. G. U. Clasificación del audio cardiaco mediante representación escasa de señales y aprendizaje automático.2019.

Khourdifi, Y., & Bahaj, M. Heart disease prediction and classification using machine learning algorithms optimized by particle swarm optimization and ant colony optimization. International Journal of Intelligent Engineering and Systems, 12(1), 242-252. 2019.

Li, J. P., Haq, A. U., Din, S. U., Khan, J., Khan, A., & Saboor, A. Heart disease identification method using machine learning classification in e-healthcare. IEEE Access, 8, 107562-107582. 2020.

Li, J. P., Haq, A. U., Din, S. U., Khan, J., Khan, A., & Saboor, A. Heart disease identification method using machine learning classification in e-healthcare. IEEE Access, 8, 107562-107582. 2020.

Llodrà Bisellach, G. “Aprendizaje automático para la clasificación de arritmias cardíacas”. 2018.

Mohan, S., Thirumalai, C., & Srivastava, G. Effective heart disease prediction using hybrid machine learning techniques. IEEE access, 7, 81542-81554.2019.

Montoya, R. A., Santa Chávez, J. J., & Mora, J. D. J. V. Aplicación del aprendizaje automático con árboles de decisión en el diagnóstico médico. Cultura del cuidado, 10(1), 63-72.2013.

Organización Mundial de la Salud. Enfermedades cardiovasculares. 2021. Disponible en: https://www.who.int/es/health-topics/cardiovascular-diseases#tab=tab_1

Pérez Soria, B. Explorando factores de riesgo de insuficiencia cardíaca a través del aprendizaje automático (Bachelor’s thesis, Universitat Politècnica de Catalunya). 2019.

Rajdhan, A., Agarwal, A., Sai, M., Ravi, D., & Ghuli, P. Heart disease prediction using machine learning. International Journal of Research and Technology, 9(04), 659-662. 2020.

Ramalingam, V. V., Dandapath, A., & Raja, M. K. Heart disease prediction using machine learning techniques: a survey. International Journal of Engineering & Technology, 7(2.8), 684-687. 2018.

Rani, P., Kumar, R., Ahmed, N. M., & Jain, A. A decision support system for heart disease prediction based upon machine learning. Journal of Reliable Intelligent Environments, 7(3), 263-275. 2021.

Zapana Calderon, R. F. (2021). Determinación de valvulopatías cardiacas a través de análisis de sonidos del corazón mediante algoritmos de machine learning. 2021.