

XIKUA Boletín Científico de la Escuela Superior de Tlahuelilpan
13º Congreso Internacional de Computación
Inteligencia artificial: Presente y futuro
Red Iberoamericana de Computación
Vol. 12, Número Especial (2024) 89-94

Predicción de enfermedades cardiovasculares utilizando redes neuronales artificiales

Cardiac disease prediction using artificial neural networks

*Manuel Alejandro Carrascal Arias^a, Alvaro Maximiliano Saiquita^b, Guillermo Sánchez Flores^c,
Torcoroma Velásquez Pérez^d*

Abstract:

Cardiovascular diseases include diseases of the heart and blood vessel system including the brain, legs and lungs. These diseases have a very high mortality rate and it is estimated that by the year 2030 there will be about 23.6 million people who may die from this cause. A quantitative research methodology with descriptive scope is used. This project deals with the study of a type of heart disease, heart failure (HF), which is characterized by the inability of the heart to pump blood in adequate quantities to meet the demands of metabolism. Specifically, the prediction of the disease was sought by means of artificial neural networks, one of the most suitable machine learning (ML) techniques for this type of activity, although not the only one. The implementation was based on two tools that incorporate the management of neural networks (NR), IBM's SPSS and Matlab. An accuracy of 94.7% was achieved. Additionally, the study was oriented to determine those characteristics with more influence in the prediction of CI. Both software were able to agree that the slope of the ST segment has the greatest impact on prediction.

Keywords:

Cardiovascular Diseases, Machine Learning, Artificial Neural Networks, Prediction, Heart Failure

Resumen:

Como enfermedades cardiovasculares se incluyen las enfermedades del corazón y del sistema de vasos sanguíneos que incluyen al cerebro, las piernas y los pulmones. Estas enfermedades tienen una tasa de mortalidad bastante alta y se estima que para el año 2030 serán cerca de 23.6 millones de personas las que pueden fallecer por dicha causa. Se utiliza una metodología de investigación cuantitativa con alcance descriptivo. Este proyecto aborda el estudio de un tipo de cardiopatía, la insuficiencia cardíaca (IC), la cual se destaca por la incapacidad del corazón de bombear sangre en las cantidades adecuadas y necesarias para satisfacer las demandas del metabolismo. Concretamente se buscó la predicción de la enfermedad por medio de redes neuronales artificiales, una de las técnicas de machine learning (ML) más adecuadas para este tipo de actividades, aunque no la única. La implementación se hizo a partir de dos herramientas que incorporan el manejo de redes neuronales (RN), SPSS de IBM y Matlab. Se consiguió una precisión del 94.7%. Adicionalmente, él estudio se orientó a determinar aquellas características con más influencia en la predicción de IC. Ambos softwares lograron coincidir en que la pendiente del segmento ST posee el mayor impacto en la predicción.

Palabras Clave:

Enfermedades cardiovasculares, Aprendizaje Automático, Redes Neuronales Artificiales, Predicción, Insuficiencia Cardíaca

^a Autor de Correspondencia, Universidad Francisco de Paula Santander Ocaña, <https://orcid.org/0009-0004-6941-3812>, Email: macarrascala@ufpso.edu.co

^b Universidad Nacional de Jujuy, <https://orcid.org/0009-0002-6643-0508>, Email: 42804342@fi.unju.edu.ar

^c Instituto Politécnico Nacional, <https://orcid.org/0009-0004-8691-8433>, Email: gsanchezf1900@alumno.ipn.mx

^d Universidad Francisco de Paula Santander Ocaña, <https://orcid.org/0000-0002-2968-2338>, Email: tvelasquezp@ufpso.edu.co

Introducción

La insuficiencia cardíaca (IC) es una enfermedad con alta prevalencia a nivel mundial asociada a una morbilidad considerable, altos costos y mal pronóstico a mediano plazo ^[1]. La IC se caracteriza por la incapacidad del corazón para bombear sangre de manera eficiente, concretamente defectos en el ventrículo izquierdo. Esta afección representa una carga significativa para los sistemas de atención médica, especialmente en países como Colombia donde se atraviesa por una crisis en los servicios de salud, donde en ocasiones los pacientes no pueden ser atendidos con el tiempo suficiente para curar las enfermedades o diagnosticarlas a tiempo ^[2]. A medida que pasan los años las enfermedades cardíacas van en aumento debido al envejecimiento de la población y los constantes cambios del estilo de vida de las personas. Es por ello que, en la búsqueda de una solución efectiva para el diagnóstico preciso de la IC, se considera que es esencial los avances en el campo de la inteligencia artificial y el aprendizaje profundo. Dicho esto, es fundamental resaltar el cómo en investigaciones previas los modelos basados en redes neuronales y aprendizaje profundo han demostrado un camino prometedor para la predicción de enfermedades cardiovasculares ^{[2][3]}.

En ^[4] Se usan RN de aprendizaje profundo para la clasificación de sonidos cardíacos en tiempo real, usando una red de regresión Softmax simple (SMR) y una RN convolucional (CNN), en dispositivos portátiles. En otra investigación se propuso predecir la enfermedad cardíaca utilizando factores multivariantes, se induce con aplicación secuencial de algoritmos de regresión logística y algoritmos de análisis discriminante lineal basados en técnicas de ML ^[3]. Para la predicción precoz de las reacciones adversas a los medicamentos se establecieron modelos que permitan predecir el deterioro clínico. Se introdujo vectorial (SVM), bosque aleatorio (RF), AdaBoost, XGBoost, y RN para afrontar el reto de la alta dimensionalidad ^[5].

Como técnicas de validación se planteó un sistema inteligente para predecir el porcentaje de grasa corporal en adolescentes basado en variables sociodemográficas y variables clínicas del sistema público de enseñanza en Brasil con un rendimiento predictivo se evaluó mediante la precisión, sensibilidad, especificidad ^[6]. De igual manera en ^[7] se propone un marco de detección de la salud cardíaca denominado FogDLearner el cual utiliza recursos distribuidos para diagnosticar la salud cardíaca de una persona utiliza un clasificador basado en aprendizaje profundo. En otra investigación se plantea la capacidad de diagnosticar enfermedades obteniendo toda la información de una ID sanitaria vinculada,

combinada con técnicas de aprendizaje automático para mejorar el diagnóstico de enfermedades ^[8].

Considerando los estudios previos y la importancia del tema ya que la IC es solo una de las variadas cardiopatías o enfermedades cardiovasculares. Este trabajo se enfoca en su predicción mediante el uso de redes neuronales artificiales (RN), esta técnica ofrece la posibilidad de mejorar significativamente la capacidad de predicción y control de la enfermedad. El propósito de la investigación es desarrollar un modelo de predicción, a partir de la identificación de factores de riesgo relevantes, la validación del modelo y la evaluación de su rendimiento. La metodología planteada es de corte cuantitativo y alcance descriptivo, se utilizan las redes neuronales artificiales para el entrenamiento del modelo y la validación se realiza evaluando la sensibilidad y especificidad por medio de la curva ROC, se utilizan herramientas como MATLAB y SPSS.

Metodología y procesos de desarrollo

Obtención y preparación de datos. El conjunto de datos que se utiliza para alimentar la red neuronal provienen de un dataset público creado en septiembre de 2021 almacenado en la plataforma Kaggle ^[9]. Este dataset es el resultado de la combinación de otros datasets independientes que contiene datos de cinco sitios diferentes: La Fundación Clínica de Cleveland, El Instituto Húngaro de Cardiología, Instituto de Cardiología de Suiza, incluyendo otro dataset llamado Stalog (Heart). El dataset seleccionado cuenta con doce atributos que pueden ser de representación numérica, binaria o nominal. Los atributos seleccionados son:

Edad (Age): edad del paciente en años. Es una variable numérica que indica la edad del paciente al momento del registro. Tipo de dolor torácico (ChestPainType): Este atributo describe el tipo de dolor torácico que el paciente experimenta y se codifica de la siguiente manera: [0: TA, 1: ATA, 2: NAP, 3: ASY]. Presión arterial en reposo (RestingBP): Este atributo representa la presión arterial del paciente en reposo que es medida en milímetros de mercurio [mm Hg]. Colesterol sérico (Cholesterol): Indica el nivel de colesterol sérico del paciente en miligramos por decilitro. [mm/dl]. Frecuencia cardíaca máxima alcanzada (MaxHR): Este atributo indica el valor máximo alcanzado de la frecuencia cardíaca del paciente durante el ejercicio [valor numérico entre 60 y 202]. Depresión del segmento ST (Oldpeak): Representa la magnitud de la depresión del segmento ST, este valor se utiliza como indicador de la actividad cardíaca durante el ejercicio [Valor numérico medido en depresión].

Se cuenta con atributos de representación binaria como:

Sexo (Sex): Indica el sexo del paciente y se representa con dos categorías. [1: Masculino, 0: Femenino]. **Azúcar en sangre en ayunas (FastingBS):** Este es un atributo binario que representa si el paciente tiene azúcar en sangre en ayunas: [1: Si el nivel de azúcar en sangre en ayunas es mayor que 120 mg/dl, 0: De ser el caso contrario]. **Resultados del electrocardiograma en reposo (RestingECG):** Este atributo describe los resultados del electrocardiograma en reposo y se codifica de la siguiente manera [0: Normal, 1: Con anomalía de la onda, 2: Anomalía de la onda ST-T, 3: Mostrando hipertrofia ventricular izquierda probable o definitiva según los criterios]. **Angina inducida por el ejercicio (Exercise Angina):** Es un atributo de representación binaria que indica si el paciente experimentó angina durante el ejercicio [1.Y: Sí, 0.N: No]. **Pendiente del segmento ST (ST Slope):** Este atributo describe la pendiente del segmento ST durante el punto más intenso del ejercicio y se codifica de la siguiente manera: [0: Pendiente ascendente, 1: Pendiente plana, 2: Pendiente descendente]. **Enfermedad cardíaca (HeartDisease):** Este atributo se emplea como un factor para predecir la presencia o ausencia de la enfermedad cardíaca en el paciente. [1: Enfermedad cardíaca, 0: normal]

Metodología

Para el desarrollo se utiliza la metodología CRISP-DM con sus fases de entendimiento del negocio, entendimiento de los datos, depuración de los datos, entrenamiento y validación del modelo. Se utiliza la red neuronal artificial para el entrenamiento del modelo y para evaluar el rendimiento de la RN se determinará el nivel de precisión logrado, se analizarán curvas ROC y áreas bajo la misma (AUC), así como matrices de confusión. Adicionalmente, se mostrarán gráficas de importancia normalizada con el fin de detectar aquellas características que incorporan mayor sensibilidad en cuanto a la predicción de la enfermedad.

Resultados

Se entrenó la red neuronal en la herramienta SPSS (Statistical Package for the Social Sciences) de IBM y posteriormente en Matlab para poder generar una discusión y análisis comparativo sobre los resultados obtenidos en cada uno de ellos. En ambos casos se efectuaron 10 ejecuciones por cada cambio en el número de neuronas en la capa oculta, que varió desde 10 hasta 18 neuronas, dejando un total de 80 ejecuciones en cada software, 160 en su totalidad.

Implementación en SPSS

Los resultados en términos de error de los modelos de RN entrenados en SPSS se presentan en la Tabla 1. Esta indica el menor error obtenido para distintos modelos que varían en cuanto al número de neuronas en la capa oculta y están basados en 10 ejecuciones en cada uno. También se explicita el error promedio de tales ejecuciones.

Tabla 1. Errores de las ejecuciones del modelo en SPSS con variaciones en la cantidad de neuronas de la capa oculta

Neuronas en la capa oculta	Error	
	Mejor	Promedio
10 Neuronas	0,07	0,1305
11 Neuronas	0,087	0,1259
12 Neuronas	0,112	0,1423
13 Neuronas	0,04	0,097
14 Neuronas	0,052	0,1012
15 Neuronas	0,057	0,0965
16 Neuronas	0,096	0,124
17 Neuronas	0,051	0,0914
18 Neuronas	0,027	0,0925

La Tabla 2 presenta los resultados de predicción obtenidos en SPSS para la RN que mejor rendimiento brindó, esta se constituyó de 18 neuronas en la capa oculta. La precisión obtenida en las etapas de entrenamiento, pruebas y reserva o validación, que en forma global obtuvieron 94.5%, 97.3% y 92.8% de precisión, respectivamente. El enfoque se asienta sobre las predicciones en los datos de prueba debido a que, a partir de las mismas, se puede determinar el rendimiento de la RN con observaciones de pacientes que no fueron vistas en el entrenamiento, evaluando su capacidad de generalización. Se puede observar de la Tabla 2 que la red, de forma global, obtiene una precisión del 94.7%.

Tabla 2. Precisión de predicción del modelo de RN de SPSS

Clasificación				
Ejemplo	Observado	Pronosticado		
		0	1	Porcentaje correcto
Entrenamiento	0	170	15	91,90%
	1	6	193	97,00%
	Porcentaje global	45,80%	54,20%	94,50%
Pruebas	0	40	0	100%
	1	2	31	93,90%
	Porcentaje global	57,50%	42,50%	97,03%
Reservas	0	36	2	94,70%
	1	3	28	90,30%
	Porcentaje global	56,50%	43,50%	92,80%
Global	0	246	17	93,50%
	1	11	252	95,80%
		49,40%	50,60%	94,70%

La curva ROC es útil para comparar diferentes modelos de clasificación y proporciona una forma de evaluar cómo se comporta un clasificador en términos de sensibilidad y especificidad en diferentes puntos de corte, dando un resultado de 0,988 (véase figura 1).

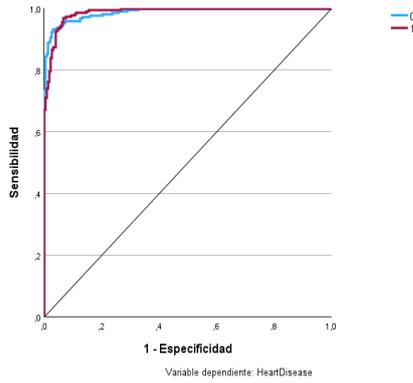


Figura 1. Curva ROC del modelo de RN de SPSS

Finalmente se presenta el gráfico de importancia normalizada (véase figura 2) en barras horizontales y en términos porcentuales, que indica la influencia relativa de cada una de las características en la determinación del objetivo, es decir, en la predicción de la enfermedad de IC. En ella se evidencia, en orden descendente de importancia, que las características pendientes del segmento ST (ST_Slope), colesterol sérico (Cholesterol) y edad (Age) son las más influyentes en la predicción de la enfermedad.

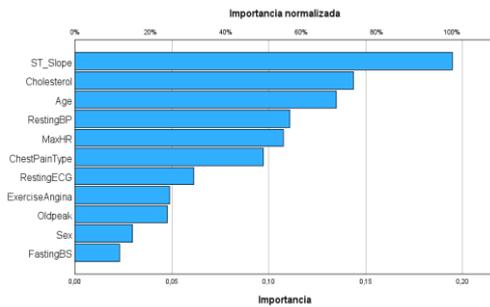


Figura 2. Gráfico de importancia normalizada del modelo de RN de SPSS

Implementación en Matlab

Para la implementación de la RN en Matlab se utilizó la toolbox de redes neuronales para el reconocimiento, clasificación de patrones y predicción. La Tabla 3 muestra los resultados de las ejecuciones variando la cantidad de neuronas en la capa oculta, en el intervalo dispuesto inicialmente. Los encabezados indican aquellas ejecuciones que obtuvieron el mejor rendimiento y un promedio de las 10 ejecuciones, ambos indicadores en términos del error cuadrático medio, que fueron de 0.0942 con 15 y 17 neuronas en la capa oculta, y 0.1507, respectivamente.

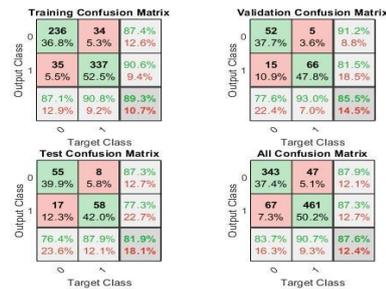
Tabla 3. Errores de las ejecuciones del modelo en Matlab con variaciones en la cantidad de neuronas de la capa oculta

Neuronas en la capa oculta	Error	
	Mejor	Promedio
10 Neuronas	0,1087	0,1507
11 Neuronas	0,1159	0,1710
12 Neuronas	0,1087	0,2583
13 Neuronas	0,1304	0,1601
14 Neuronas	0,1304	0,1529
15 Neuronas	0,0942	0,2323
16 Neuronas	0,1232	0,1616
17 Neuronas	0,0942	0,1600
18 Neuronas	0,1232	0,2590

Al igual que con las gráficas brindadas por SPSS, las gráficas a continuación están basadas en la arquitectura de RN que obtuvo el mejor rendimiento en las ejecuciones llevadas a cabo, es decir, aquella conformada por 15 neuronas en la capa oculta.

En la matriz de confusión del modelo (véase figura 3) se muestra que, para el conjunto de testeo o prueba, se obtuvo una predicción correcta de la enfermedad del 81.9% de los casos, siendo el 18.1% predicciones erróneas. Luego, en términos globales, la RN predijo correctamente el 87.6% de los casos, determinando incorrectamente la enfermedad en el 12.4% restante.

Figura 3. Matriz de confusión del modelo de RN de Matlab



En la curva ROC se puede apreciar una tendencia de la curva hacia la esquina superior izquierda, tal y como sucedía con la curva ROC dada por SPSS, alejado de la diagonal aleatoria y evidenciando nuevamente que no se trata de un modelo sin capacidad de predicción. Además, esto se comprueba evaluando el área bajo esta curva que, en este caso, es de 0.924.

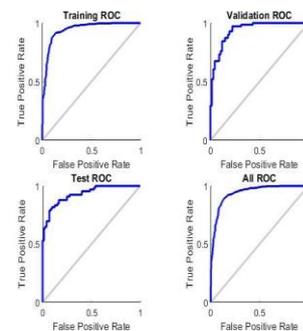


Figura 4. Curva ROC del modelo de RN de Matlab

Finalmente, se proporciona el gráfico de importancia normalizada dado en la Figura 5. Este indica que las características más influyentes en la predicción de IC son, en orden descendente de importancia, pendiente del segmento ST (ST_Slope), tipo de dolor torácico (ChestPainType) y depresión del segmento ST (Oldpeak).

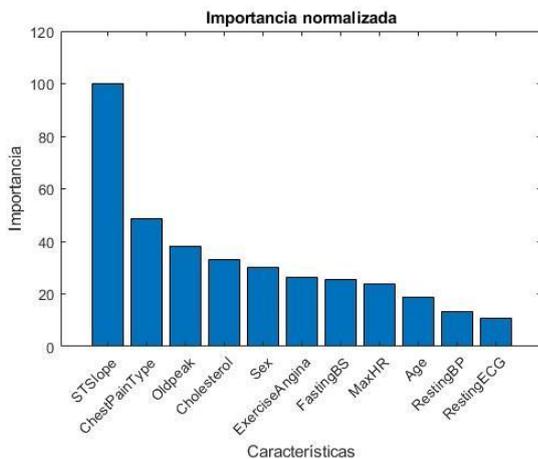


Figura 5. Gráfico de importancia normalizada del modelo de RN de Matlab

Discusión (análisis comparativo)

Las implementaciones de RN para la predicción de la enfermedad de IC en SPSS y Matlab han proporcionado resultados correctos para el ámbito en estudio, pero existe una diferencia esclarecedora.

Partiendo por las precisiones arribadas en cada caso, en términos del conjunto de prueba, el cual permite apreciar la capacidad de generalización de ambos modelos, SPSS obtuvo un 97.3% de pronósticos acertados, por sobre el 81.9% de casos correctos para Matlab. Denotando con ello una diferencia de más de 15 puntos porcentuales. En términos globales, el modelo de RN de SPSS logró una precisión del 94.7%, mientras que Matlab 87.6%. En este caso la diferencia disminuye a 7.1 puntos porcentuales.

La curva ROC de ambos modelos de RN indican un rendimiento óptimo de la red, pero destaca la curva dada por SPSS en donde esta se orienta más en la dirección superior izquierda, denotando su capacidad predictiva al disponer de más casos verdaderos positivos por sobre falsos positivos, el área bajo la curva soporta tal capacidad, siendo de 0.988 en SPSS y 0.924 en Matlab. En definitiva, puede predecir correctamente más casos reales de IC por sobre errar en la predicción o diagnóstico.

La gráfica de importancia normalizada hace coincidir a la variable o característica pendiente del segmento ST (ST_Slope) del paciente como la más influyente a la hora

de determinar el diagnóstico. Cabe recordar que este es un parámetro electrocardiográfico medido en el ejercicio máximo del paciente y su valor puede dar indicios de algún tipo de cardiopatía. Las características posteriores poseen un nivel de importancia que se diferencian en ambos modelos. SPSS dispone a la edad (Age) y el colesterol sérico (Cholesterol) como las más sensibles después de la característica pendiente del segmento ST. Matlab, por su parte, determina este peso en la influencia de las características al tipo de dolor torácico (ChestPainType) y depresión del segmento ST (Oldpeak), luego de la pendiente del segmento ST.

Es evidente que SPSS ha logrado obtener un modelo de RN considerablemente más eficiente que el arribado en Matlab. A pesar de esto, la configuración automática de parámetros de la RN en SPSS proporciona resultados similares o peores que Matlab. En su lugar, se tomó la configuración recomendada por este último para tareas de clasificación y predicción, y se migraron a SPSS, esto referido al algoritmo de entrenamiento y a las funciones de activación en la capa oculta y de salida. Por su parte, en Matlab se usó la toolbox de RN por defecto, apropiándonos de la configuración recomendada para la tarea de predicción de IC, aunque esta permitía la exportación del código en un script y con ello, la posibilidad de adentrarse en el refinamiento del modelo, aportando mayor flexibilidad que SPSS.

Conclusiones

A lo largo de esta investigación se abordó la tarea de predecir la IC utilizando modelos de RN. El enfoque no solo se centró en la identificación de la presencia o ausencia de esta afección, sino también en la identificación de las principales causas subyacentes que contribuyen significativamente al desarrollo de la enfermedad.

En un análisis inicial, los resultados que se obtuvieron respaldan la efectividad de los modelos de RN para la predicción de la IC, demostrando un alto grado de precisión y capacidad de generalización, concretamente del 94.7%. Esto sugiere su utilidad como herramienta potencialmente beneficiosa para el ámbito médico en la detección de esta enfermedad cardíaca que se considera crítica, por sobre los intereses individuales de expertos en el área.

Adicionalmente, se identificó que la pendiente del segmento ST (ST_Slope) es la característica más relevante en cuanto a la predicción de la enfermedad de IC, más no se puede concluir que sea el factor de riesgo determinante ni mucho menos la única que tenga

influencia en el objetivo, por muchas razones (tamaño y origen de las observaciones, cantidad y tipo de características de pacientes asociadas a la cardiopatía en estudio, interpretación médica de las mismas). Las características posteriores a la pendiente del segmento ST según el nivel de importancia difirieron de modelo en modelo.

Se determinó que tanto Matlab como SPSS son adecuados para la implementación de modelos de este tipo, y la elección de su uso no debe basarse en los resultados obtenidos en esta investigación. Más bien, se sugiere indagar en ambas herramientas e inclusive probar otras técnicas de que destacan para este tipo de problemáticas sanitarias y demás, lo cual será el punto de partida de nuestras próximas investigaciones.

Finalmente se señala que, aunque en esta fase inicial de la investigación se utiliza un conjunto de datos público, nuestros futuros esfuerzos estarán enfocados en ampliar este estudio en busca de una comprensión más detallada y completa de las causas y factores de riesgo específicos para la IC en México, Argentina y Colombia, utilizando instrumentos que nos permitan recopilar los datos necesarios y apoyándonos en recomendaciones y aportes de expertos en el campo de la cardiología, con el fin de lograr resultados aún más fiables y fundamentados.

Agradecimientos

A Kaggle y a los contribuyentes por los dataset. A los autores referenciados. A CICOM y a la Universidad Francisco De Paula Santander Ocaña por el evento y el apoyo. También agradecemos al Instituto Politécnico Nacional y a la Universidad Nacional de Jujuy.

Referencias

- [1] Gambarte MJ, Higa C, Novo F, Ciabrone GM, Villagomez ODT, Ginesi A, et al. Comparación pronóstica entre scores de riesgo y la aplicación de redes neuronales para la predicción de la mortalidad a corto y mediano plazo en pacientes con insuficiencia cardíaca. REVISTA ARGENTINA DE CARDIOLOGÍA. 2021;89.
- [2] Avellaneda González JA, Ochoa Rey CM. IMPLEMENTACIÓN DE REDES NEURONALES PARA LA DETECCIÓN DE LA PRESENCIA DE ENFERMEDADES EN EL CORAZÓN. redes ing. 23 de mayo de 2012;1(2):38.
- [3] Rajalakshmi S, Madhav KV, Abhishek R. Inducement of Multivariate factors in Cardiac Disease Prediction with Machine Learning Techniques substantiated with Analytics. En: Proceedings of the 2019 11th International Conference on Computer and Automation Engineering [Internet]. Perth WN Australia: ACM; 2019 [citado 10 de septiembre de 2023]. p. 97-101. Disponible en: <https://dl.acm.org/doi/10.1145/3313991.3314014>
- [4] Low JX, Choo KW. Classification of Heart Sounds Using Softmax Regression and Convolutional Neural Network. En: Proceedings of the 2018 International Conference on Communication Engineering and Technology [Internet]. Singapore Singapore: ACM; 2018 [citado 10 de septiembre de 2023]. p. 18-21. Disponible en: <https://dl.acm.org/doi/10.1145/3194244.3194255>
- [5] Zhao YX, Yuan H, Wu Y. Prediction of Adverse Drug Reaction using Machine Learning and Deep Learning Based on an Imbalanced Electronic Medical Records Dataset. En: 2021 5th International Conference on Medical and Health Informatics [Internet]. Kyoto Japan: ACM; 2021 [citado 10 de septiembre de 2023]. p. 17-21. Disponible en: <https://dl.acm.org/doi/10.1145/3472813.3472817>
- [6] Araujo C, M A Azoubel L, K D B Filho A, M S Junior C, P S Sousa N, Pereira C, et al. Prediction of Excess Body Fat in Adolescents Using Neural Networks. En: 2021 8th International Conference on Bioinformatics Research and Applications [Internet]. Berlin Germany: ACM; 2021 [citado 10 de septiembre de 2023]. p. 63-9. Disponible en: <https://dl.acm.org/doi/10.1145/3487027.3487037>
- [7] Iftikhar S, Golec M, Chowdhury D, Gill SS, Uhlig S. FogDLearner: A Deep Learning-based Cardiac Health Diagnosis Framework using Fog Computing. En: Australasian Computer Science Week 2022 [Internet]. Brisbane Australia: ACM; 2022 [citado 10 de septiembre de 2023]. p. 136-44. Disponible en: <https://dl.acm.org/doi/10.1145/3511616.3513108>
- [8] Naik KT, Garg B. A Machine Learning Model for Disease Prediction and Remote Patient Monitoring. En: Proceedings of the 4th International Conference on Information Management & Machine Intelligence [Internet]. Jaipur India: ACM; 2022 [citado 10 de septiembre de 2023]. p. 1-5. Disponible en: <https://dl.acm.org/doi/10.1145/3590837.3590844>
- [9] Fedesoriano. Heart Failure Prediction Dataset [Internet]. Septiembre 2021 [citado 10 de septiembre de 2023]. Disponible en: <https://www.kaggle.com/fedesoriano/heart-failure-prediction>.