universidad de león

FACULTAD DE CIENCIAS BIOLÓGICAS Y AMBIENTALES

COMPARACIÓN DE ALGORITMOS DE INTELIGENCIA ARTIFICIAL PARA LA PREDICCIÓN DE ADENOCARCINOMA PANCREÁTICO

COMPARING ARTIFICIAL INTELLIGENCE ALGORITHMS FOR PREDICTING PANCREATIC DUCTAL ADENOCARCINOMA

Autor: Isabel Rodríguez Valle

Tutor: María Montserrat López Cabeceira

GRADO EN BIOTECNOLOGÍA

Julio, 2022

ÍNDICE

RESUMEN

La dificultad en el diagnóstico temprano del adenocarcinoma pancreático ductal es una de las principales razones de su alta tasa de mortalidad. Con el fin de favorecer este diagnóstico se han desarrollado distintas versiones de tres tipos de algoritmos de inteligencia artificial y aproximación de datos: algoritmo genético, redes neuronales y regresión logística. Estos toman los datos de sexo, edad, CA19-9 en sangre y niveles de creatinina, TFF1, REG1B y LYVE1 en orina de cientos de pacientes y los clasifican como casos control, los que presentan un tumor benigno o los que padecen uno maligno. Se ha estudiado la precisión, especificidad y sensibilidad de estos algoritmos, determinando cuáles conllevan una mayor capacidad de predicción, y comparando estos entre sí y con los resultados obtenidos en bibliografía anterior, siendo la novedad el algoritmo genético. La capacidad predictiva de este ha resultado ser comparable a la de los ya estudiados, obteniendo la mayor precisión para la clasificación entre controles y tumores malignos. Además, se obtienen sensibilidades y especificidades mayores al 80% para los tres métodos para esta clasificación. Esto confirma el potencial de las herramientas de *machine learning* para el diagnóstico de este tipo de tumor, aunque aún existan limitaciones para su implantación clínica.

ABSTRACT

The low rate of early-stage diagnosis of pancreatic ductal adenocarcinoma is one of the main reasons for its high mortality rate. In order to benefit this diagnosis, different versions of three types of artificial intelligence and data approximation algorithms have been developed: genetic algorithm, neural networks and logistic regression. These take collected data on sex, age, CA19-9 in blood and levels of creatinine, TFF1, REG1B and LYVE1 in urine from hundreds of patients and classify them as control cases, those with a benign tumor and those with a malignant one. The accuracy, specificity and sensitivity of these algorithms have been studied, determining which ones entail a greater prediction capacity, and comparing them with each other and with the results obtained in previous research, the novelty being the genetic algorithm. This algorithm's predictive capacity has turned out to be comparable to that of those already studied, resulting in the highest accuracy for classification between controls and malignant tumors. Furthermore, sensitivities and specificities greater than 80% have been achieved for the three methods for this classification. This confirms the potential of machine learning tools for diagnosing this type of tumor, although there are still limitations to its clinical implementation.

Palabras clave: Adenocarcinoma pancreático ductal, Algoritmo genético, Diagnóstico clínico, Inteligencia artificial, Red neuronal, Regresión logística.

Key words: Artificial intelligence, Clinical diagnosis, Genetic algorithm, Logistic regression, Neural network, Pancreatic ductal adenocarcinoma.

1. INTRODUCCIÓN

El adenocarcinoma pancreático ductal (*pancreatic ductal adenocarcinoma* o PDAC) es la clase de tumor maligno más prevalente en el páncreas y conlleva una alta gravedad médica (Collisson *et al*., 2019). Tiene una tasa de supervivencia a 5 años de tan solo el 11% (Siegel *et al*., 2022), con más del 90% de pacientes falleciendo antes de haber pasado 1 año de su diagnóstico (Stott *et al*., 2022). El desafortunado pronóstico se atribuye al desarrollo biológico de la enfermedad, la falta de pautas internacionales para el estudio de masas sospechosas y a la dificultad para obtener un diagnóstico temprano, es más, suele ser diagnosticado cuando el cáncer ya se considera metastásico (Zhang *et al*., 2018).

Esta tardanza se debe a que los síntomas descritos no son específicos, incluyendo pérdida de peso, dolor de espalda o abdominal, náuseas y vómitos, diarrea o estreñimiento y aparición de diabetes, o ictericia en pacientes más jóvenes (Loveday *et al*., 2019). Además, una gran cantidad de casos son asintomáticos (Zhang *et al*., 2018). También influye la cercanía del tumor a los principales vasos sanguíneos, lo que facilita la invasión. Todo esto contribuye a que entre el 80 y el 85% de tumores no sean extirpables cuando se encuentran, siendo la extracción quirúrgica la única cura disponible en la actualidad, aunque la probabilidad de supervivencia se mantenga baja (McGuigan *et al*., 2018). Se piensa que para poder conseguir esta curación se debe detectar el tumor cuando su tamaño es menor a 1 cm (Wada *et al*., 2015).

Se puede concluir que un diagnóstico temprano es clave cuando se trata de PDAC, pero esto no pretende conseguirse mediante un cribado de toda la población, ya que no es factible ni rentable. Se realizará tan solo en aquellos individuos considerados de alto riesgo, incluyendo a aquellos que tengan al menos dos parientes cercanos con PDAC, los que padezcan de ciertos síndromes genéticos como el de Peutz-Jegher o mutaciones en la línea germinal en el gen *CDKN2A* o en los genes *BRCA2, BRCA1, PALB2, ATM, MLH1, MSH2,* o *MSH6* si además tienen un pariente afectado. Por último, se recomienda también en aquellos con lesiones quísticas mucinosas en el páncreas (Pereira *et al*., 2020; Del Chiaro *et al*., 2013; Goggins *et al*., 2020).

Los métodos de cribado incluyen imagen por resonancia magnética, colangiopancreatografía por resonancia magnética, aspiración con aguja fina, tomografía computarizada, tomografía por emisión de positrones y ultrasonido endoscópico, siendo este último el que presenta una mayor sensibilidad (Wiest *et al*., 2020).

A pesar de esto, todos estos métodos convencionales no parecen resultar en una sensibilidad ni especificidad suficiente para el diagnóstico de lesiones antes de volverse malignas. El estudio de biomarcadores se presenta como la mejor alternativa a los mismos, permitiendo estudiar fluidos corporales de forma poco invasiva, más barata y permitiendo un diagnóstico en fases anteriores (Brezgyte *et al.*, 2021).

El antígeno carbohidrato 19-9 (CA19-9) indica glicosilación aberrante de proteínas que puede acelerar la progresión del cáncer pancreático, además de promover la angiogénesis y mediar la respuesta inmune (Luo *et al*., 2021). Es el biomarcador más popular y el único marcador sérico aprobado por la *Food and Drug Administration* (FDA) de Estados Unidos, pero no se cree que este tenga suficiente especificidad para ser usado en cribados extensivos. A pesar de esto, sí se ha probado su utilidad como biomarcador de pronóstico, pudiendo ayudar a predecir la supervivencia de un paciente tras someterse a tratamiento (Yang *et al*., 2021; Poruk *et al*., 2013).

Se han realizado cientos de estudios con el objetivo de encontrar un biomarcador que pueda facilitar el diagnóstico, ya fueran microRNAs, proteínas, exosomas o células DNA tumoral circulante provenientes de diversos medios (saliva, orina, heces, suero, jugo pancreático o fluido quístico) pero no se ha dado con uno idóneo. A pesar de esto, se ha podido comprobar que los paneles de biomarcadores son más prometedores que el estudio de uno en concreto. También se pueden afirmar las ventajas de usar métodos menos invasivos, como la orina o la saliva (O'Neill y Stoita, 2021).

Teniendo esto presente, se ha estudiado el proteoma urinario de pacientes sanos como control, con pancreatitis crónica (PC) como condición benigna, y con PDAC. Se determinó que las siguientes proteínas presentaban niveles de expresión elevados en los tumores malignos: receptor de hialuronano endotelial de los vasos linfáticos 1 (LYVE1), proteína regeneradora derivada del islote de Langerhans 1 alfa (REG1A) y 1 beta (REG1B), y factor trefoil 1 (TFF1) (Fig. 1), siendo prometedoras para su uso como panel de biomarcadores para diagnóstico apoyadas por CA19-9 plasmático (Radon *et al*., 2015).

LYVE1 se une a las formas soluble y no soluble del ácido hialurónico y se cree que puede participar en el transporte linfático del mismo, favoreciendo la metástasis tumoral y actuando como indicador de linfangiogénesis (Jackson, 2003; Li *et al*., 2018).

REG1A y REG1B pertenecen a la familia de proteínas regeneradoras "REG", compuesta por cinco proteínas similares a la lectina de tipo C. Se ha visto que la sobreexpresión del gen de REG1A lleva a proliferación celular y crecimiento tumoral. Además, ambas REG están presentes en concentraciones altamente significativas en pacientes tanto con PDAC cuando se

comparan con pacientes control. No sólo entonces, sino que ya aparecen en los análisis séricos en cantidades cada vez mayores a medida que progresa la neoplasia intraepitelial pancreática (PanIN), precursora del PDAC (Li *et al*., 2016).

TFF1 pertenece a una familia de péptidos expresados y secretados en la mucosa gastrointestinal. Normalmente, protegen las células epiteliales de la apoptosis y aumentan su motilidad, pero también se ha visto que están involucradas en el desarrollo de distintos tipos de cáncer. Se ha podido determinar que el aumento de la expresión de TFF1 está altamente relacionado con las etapas iniciales de PanIN, pero que esta expresión se pierde cuando se convierte en PDAC y desarrolla invasividad, pudiendo encontrar mayores cantidades en el centro del tumor (Klett *et al.*, 2018; Yamaguchi *et al*., 2018).

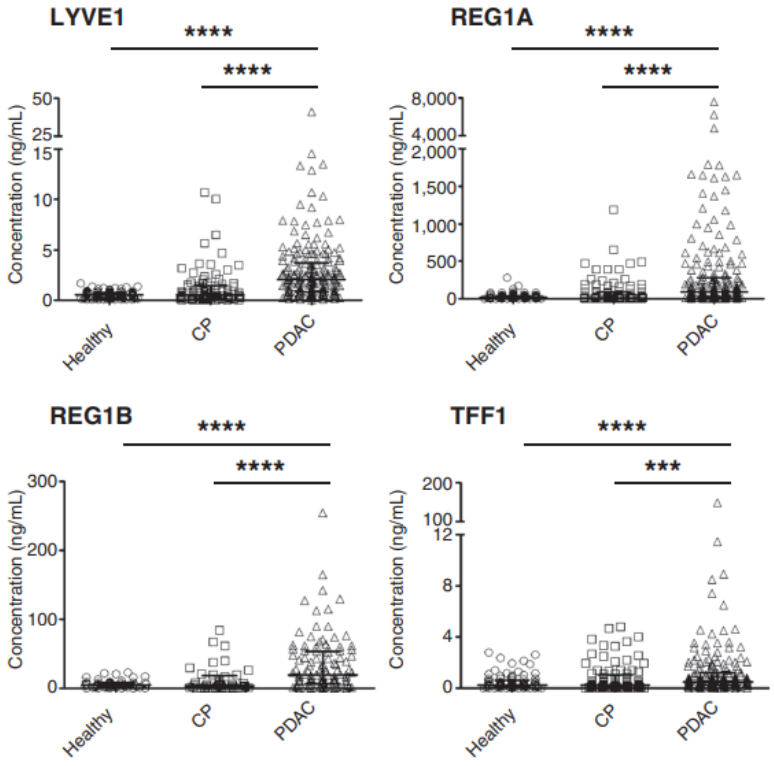


Figura 1. Diagrama de puntos que representa la concentración de los biomarcadores analizados (LYVE1, REG1A, TFF1 y REG1B) en pacientes sanos, con PC y con PDAC. Las concentraciones han sido normalizadas con respecto a la creatinina y han sido calculadas mediante un ensayo por inmunoabsorción ligado a enzimas (ELISA). Las barras superiores indican el resultado del test de Kruskal–Wallis (Kruskal y Wallis, 1952); **** , P< 0.0001; *** , P <0.001 Imagen tomada de Radon *et al*., 2015.

Para realizar el diagnóstico a partir de estos biomarcadores, se necesita un algoritmo que ayude a discriminar el estado del paciente mediante los niveles proteicos medidos en su orina, basándose en diagnósticos confirmados, obtenidos previamente por otros métodos.

En esto se basan las técnicas de inteligencia artificial (IA) y *machine learning* (ML). Se utilizan para identificar agrupaciones de datos que pueden aportar información de interés o

para el ajuste de modelos predictivos a datos obtenidos, como en este caso; tratando de imitar la capacidad de reconocimiento de patrones humana pero con una mayor eficiencia cuando los datos son demasiados o muy complejos, como suele ocurrir al hablar de datos biomédicos (Greener *et al*., 2022). Estas técnicas son muy flexibles, incluyendo entre sus utilidades en el campo de la medicina la estratificación de riesgos, la clasificación, la predicción de supervivencia o el diagnóstico, pudiendo usar como entradas distintos tipos de datos, desde imágenes y datos genómicos a notas tomadas por los practicantes (Ngiam y Khor, 2019).

En el campo de la oncología, ML es una herramienta muy prometedora con la que se espera poder mejorar significativamente el proceso de toma de decisiones de los profesionales médicos, proporcionando diagnósticos de forma más rápida, barata y precisa. No sólo son útiles de por sí, sino que además pueden usarse para complementar la percepción humana, por ejemplo, se ha estudiado que un algoritmo de ML ayudó a reducir la tasa de error en la identificación de cáncer de pecho metastásico de un 3% a un 1% (Tseng *et al*., 2020). ML se ha utilizado para el estudio de diversas neoplasias, entre otras, para la predicción de la progresión de cáncer en las cavidades orales (Adeoye *et al*., 2021), de la supervivencia de pacientes que padecen cáncer de pulmón (Lynch *et al.*, 2017) o más recientemente para predecir el pronóstico de pacientes con cáncer de colon no metastásico (Tang *et al*., 2022). A pesar de la gran cantidad de estudios existentes, a fecha de abril de 2021, tan solo 49 dispositivos han sido aprobados por la FDA para su uso clínico, la mayoría con el fin de diagnosticar, con 8 de los mismos pensados para el cáncer de mama y ninguno para el cáncer pancreático (Lyell *et al.*, 2021).

Los pacientes con PDAC resultan ser uno de los grupos más difíciles de estratificar (Hu *et al*., 2019), pero existen distintos enfoques y áreas de investigación cuyo objetivo es poder diagnosticarlo lo antes posible. Estos enfoques incluyen el uso de imágenes, provenientes, por ejemplo, de tomografías computarizadas abdominales (Wang *et al*., 2018) o ensayos de biomarcadores en sangre como CancerSEEK, que analiza niveles de proteínas y mutaciones en el DNA libre celular (Cohen *et al*., 2018). También puede utilizarse el estudio del genoma, los factores inmunológicos y el microbioma del paciente, así como la información obtenida de tests clínicos, su historial clínico o estilo de vida, o incluso aquellos datos que se pueden encontrar en sus redes sociales (Kenner *et al*., 2021).

Para el presente estudio, se utilizarán, como ya se ha mencionado anteriormente, datos de las proteínas buscadas presentes en la orina de los pacientes, así como los pertenecientes a niveles de CA19-9.

Blyuss y colaboradores desarrollaron un algoritmo llamado PancRISK, basado en las concentraciones de TFF1, LYVE1 y REG1B, así como en la edad y los niveles de creatinina. Para ello, probaron distintos tipos de ML para determinar qué modelo permitiría una más pronta detección de PDAC: máquinas de vector soporte (SVM), redes neuronales artificiales (NN), estas redes combinadas con lógica difusa (NFT), *random forest* (RF) y regresión logística (RL). Como puede verse en la Figura 2, se compararon los algoritmos, y se determinó mediante el test de McNemar que ninguno era significativamente mejor a la RL, que se trata del más sencillo de interpretar, por lo que este fue el elegido para PancRISK (Blyuss *et al.*, 2020).

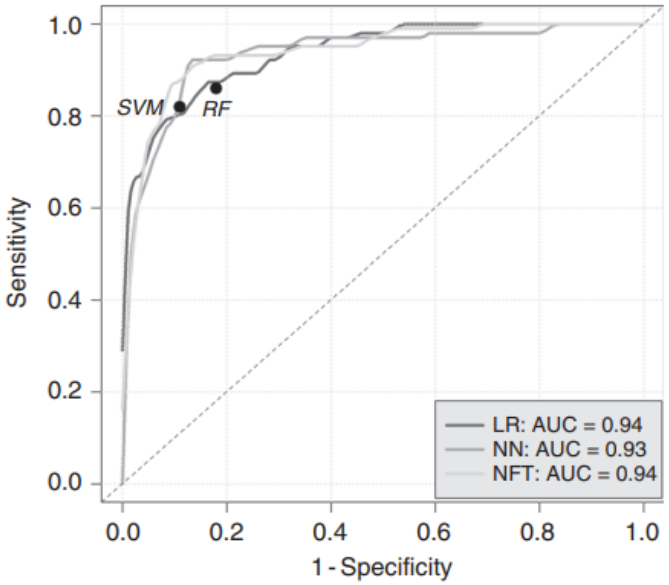


Figura 2. Rendimiento de los distintos algoritmos representado en una gráfica de *Receiver Operating Characteristic* (ROC) para LR, NN y NFT y como valores puntuales para SVM y RF. Se indica también el área bajo la curva (AUC) de ROC. Gráfica tomada de Blyuss *et al*., 2020.

2. OBJETIVO Y PROCEDIMIENTO

El objetivo de este estudio consiste en la comparación de la capacidad de predicción de los métodos NN y RL usados durante el desarrollo de PancRISK con la de un tipo de algoritmo no considerado en ese estudio (Blyuss *et al*., 2020), un algoritmo genético (AG). Su rendimiento se compara además con los resultados publicados en Debernardi y colaboradores (2020) al aplicar PancRISK para el mismo conjunto de datos usados en el presente trabajo.

Para alcanzar tal objetivo, la metodología se basa en una serie de pasos. Primero, se realiza una búsqueda bibliográfica, con el fin de obtener un conjunto de datos ya publicado, así como para encontrar artículos con objetivos similares que sirvan como precedentes teóricos. Los datos usados se obtienen de la plataforma Kaggle (Davis, 2020), en la que se ha hecho público el conjunto de datos de los pacientes estudiados por Debernardi y colaboradores en su artículo "*A combination of urinary biomarker panel and PancRISK score for earlier detection of pancreatic cancer: A case–control study*" (Debernardi *et al.*, 2020).

A continuación, se procesan los datos obtenidos para su uso, convirtiendo los valores deseados a notación numérica y eliminando los valores atípicos y aquellas muestras de las que falte información. Para este procesamiento, así como para el desarrollo de los algoritmos, se utiliza el entorno de desarrollo integrado (IDE) RStudio (RStudio Team, 2022) y el lenguaje de programación Python. Para la realización de los diagramas se ha utilizado el programa Dia (The Dia Developers, 2014) y la herramienta de dibujo de Docs.google.com.

Posteriormente, se efectúan numerosas pruebas con los programas realizados, variando las fórmulas utilizadas con el fin de obtener la mayor especificidad y sensibilidad posible para cada tipo de algoritmo. Las distintas versiones y los distintos tipos de clasificación se comparan entre sí, y también contra los resultados obtenidos en el artículo de partida (Debernardi *et al*., 2020).

Para todo el proceso se utilizan los siguientes paquetes y librerías:

- reticulate (Ushey *et al*., 2022): Interfaz para utilizar Python dentro de una sesión de R, permitiendo interoperabilidad y la importación de módulos de Python, y traduciendo entre objetos de ambos lenguajes.
- NumPy (Harris *et al*., 2020): Librería de Python esencial para la computación científica, facilitando operaciones con *arrays*, incluyendo funciones matemáticas, logísticas, estadísticas, de simulación aleatoria, etc.
- pandas (Reback *et al*., 2022): Librería de Python que proporciona herramientas para el análisis de datos de alta eficiencia.
- SciPy (Virtanen *et al*., 2020): Software con herramientas para uso científico y matemático en Python.
- math (Van Rossum, 2020): Módulo de Python que proporciona acceso a funciones matemáticas, usado principalmente para el procesamiento inicial de datos.
- ROCR (Sing *et al*., 2005): Paquete de R utilizado para la visualización de mediciones de rendimiento como las gráficas ROC o las curvas de sensibilidad/especificidad.

3. DESARROLLO DEL PLANTEAMIENTO

Para comenzar, se modifican los datos publicados en Debernardi y colaboradores (2020) para poder trabajar con ellos, lo que será común para todas las pruebas. Primero, se seleccionan las columnas (atributos) con las que se va a trabajar, aquellas que recogen los datos de sexo, edad, CA19-9 en plasma, creatinina, LYVE1, REG1B, TFF1 y diagnóstico. Se descarta la columna de REG1A por la alta cantidad de datos perdidos, así como por haberse determinado que su valor predictivo es menor al de REG1B (Blyuss *et al.*, 2020).

Después, se eliminan aquellas filas (muestras) que contengan valores señalados como *not a number* (*NaN*), que indican que el dato al que corresponde la columna no está disponible para esa muestra, ya que se leerían como 0, impidiendo el correcto funcionamiento del programa. A continuación, se convierten los valores del sexo en numéricos y se seleccionan las columnas que contienen entradas, es decir, todas menos el diagnóstico que constituye la salida esperada. Se eliminan aquellas filas que tengan una puntuación z mayor de 4 y, posteriormente, se normalizan los valores entre 0 y 1. La puntuación z o puntuación estándar permite calcular cuántas unidades de desviación estándar se aleja cada valor en el conjunto de datos de la media de su columna, permitiendo descartar los valores atípicos.

Tras esto, se convierte el diagnóstico de forma diferente según el planteamiento del programa. Este está escrito originalmente como 1 para pacientes sanos, 2 para aquellos con PC, considerado un tumor benigno, y 3 para PDAC o maligno.

Por último, se divide el conjunto de datos como 80% para entrenamiento y 20% para validación, pero estos porcentajes varían en cada prueba, ya que tras la división se iguala el número de muestras para cada diagnóstico en el conjunto de entrenamiento, con el fin de evitar parcialidad, eliminando los casos sobrantes de este conjunto y pasándolos al de validación.

Véase el Anexo I para la implementación de algunas partes del código, incluyendo este procesamiento de los datos (Fig. 15).

3.1 ALGORITMO GENÉTICO

Los AG son herramientas de optimización que llevan siendo populares desde hace más de 30 años (Whitley, 2019). Resultan especialmente útiles en casos de optimizaciones con más de un objetivo en las que se necesita reducir las dimensiones de las muestras e incrementar la precisión de clasificación (Ghosh *et al*., 2019). Se basan en las leyes de la genética para buscar posibles soluciones, que comienzan como un conjunto de individuos aleatorios con

distintas propiedades que serían sus cromosomas (Ghaheri *et al*., 2015). Para producir la siguiente generación, incrementando la diversidad, se utilizan distintos operadores de codificación, selección, cruce o *crossover* y mutación (Fig. 3).

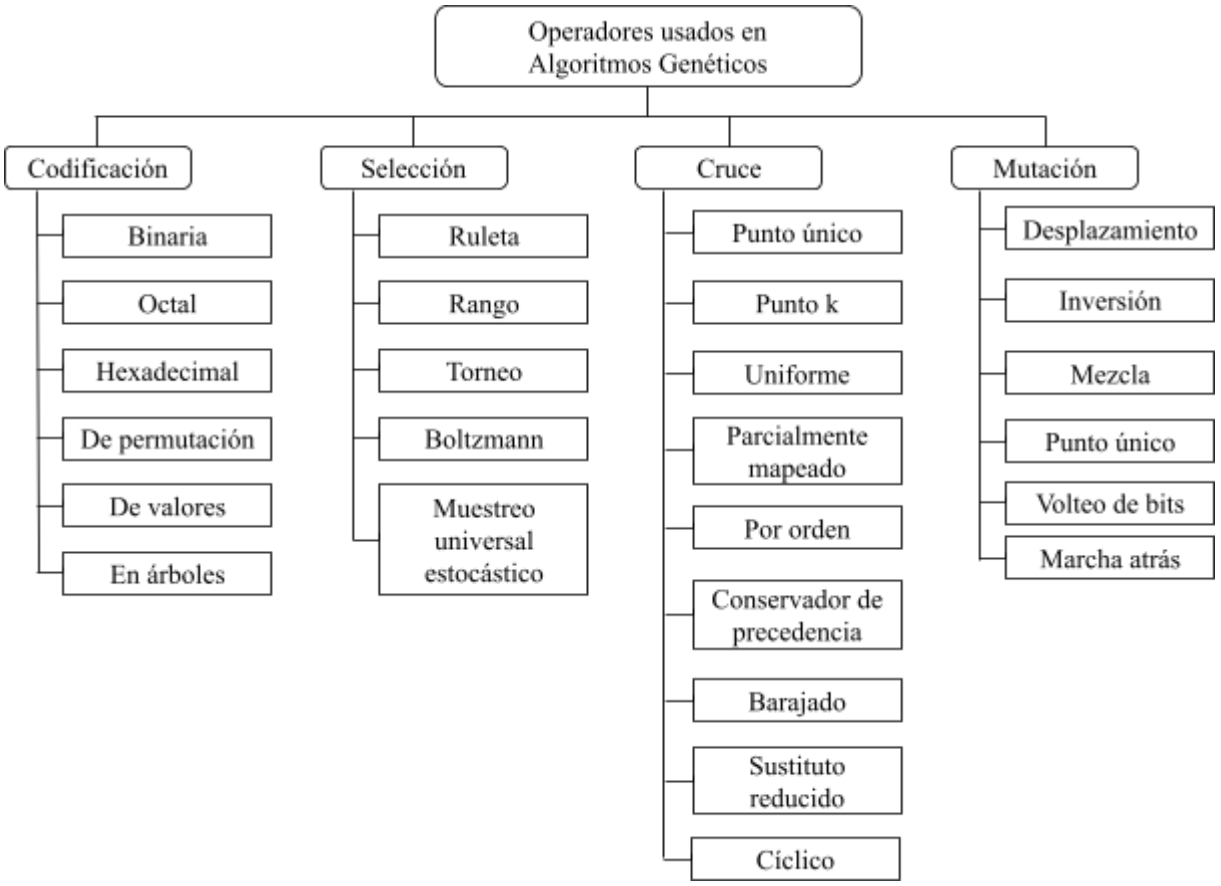


Figura 3. Posibles esquemas de codificación, técnicas de selección del mejor individuo, métodos de *crossover* y tipos de mutación existentes para aplicar en un AG. Esquema adaptado de Katoch *et al*., 2021.

El objetivo del AG escrito es encontrar, mediante una función de aptitud idónea, un individuo que nos permita diagnosticar a un paciente al aplicar la función de aptitud sobre su muestra, estando esta compuesta de los parámetros medidos. Para ello seguirá los pasos descritos en la Figura 4. Se eligen como operadores la codificación por valores, ya que se trabaja con los datos como números enteros, la selección mediante el método de la ruleta, en la que la probabilidad de un individuo de ser seleccionado es proporcional a su aptitud (Fig. 5), el *crossover* de punto único (Fig. 6) y para la mutación se cambia el valor a mutar dentro del cromosoma por otro aleatorio que se encuentra en un rango mayor al rango usado para crear la población inicial.

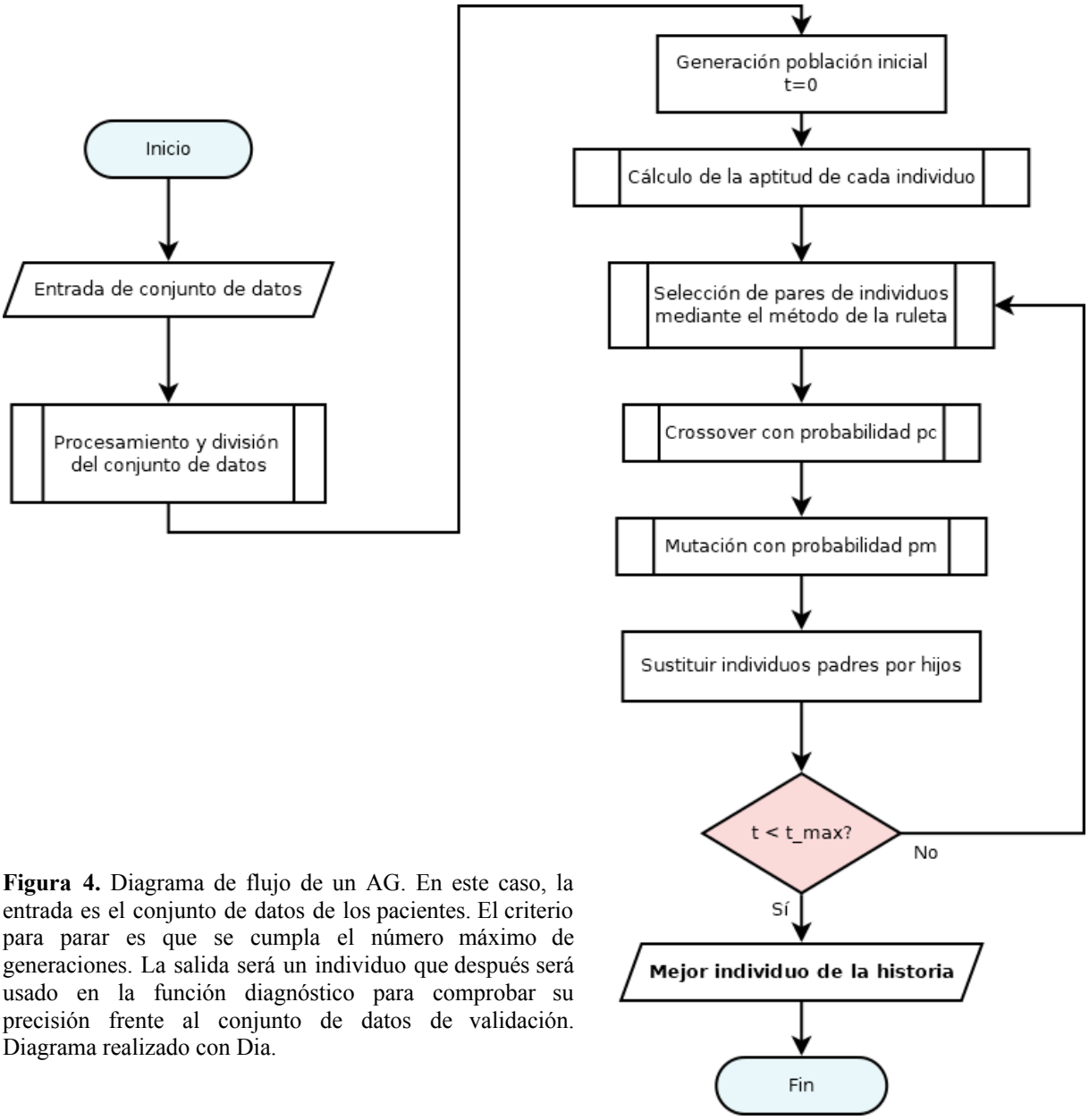
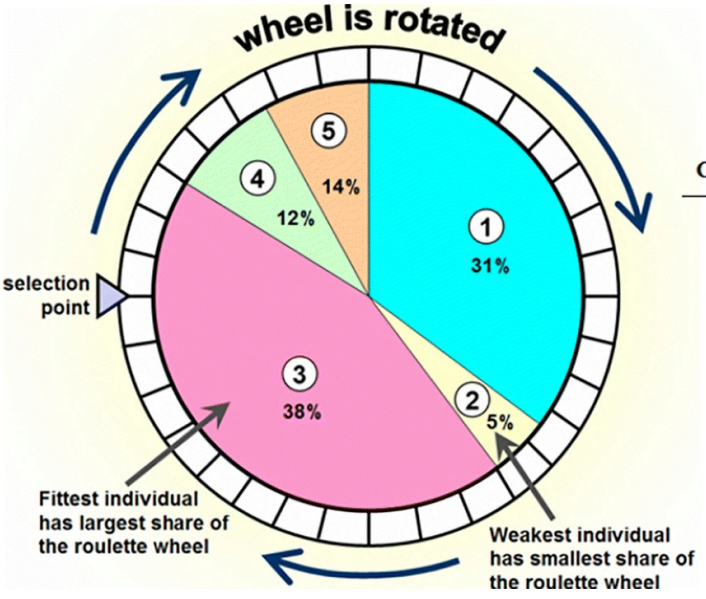


Figura 4. Diagrama de flujo de un AG. En este caso, la entrada es el conjunto de datos de los pacientes. El criterio para parar es que se cumpla el número máximo de generaciones. La salida será un individuo que después será usado en la función diagnóstico para comprobar su precisión frente al conjunto de datos de validación. Diagrama realizado con Dia.



Chromosome	Fitness value
C1	31%
C2	5%
C3	38%
C4	12%
C5	14%

Figura 5. Selección por el método de la ruleta. Figura tomada de Faradonbeh *et al.*, 2018.

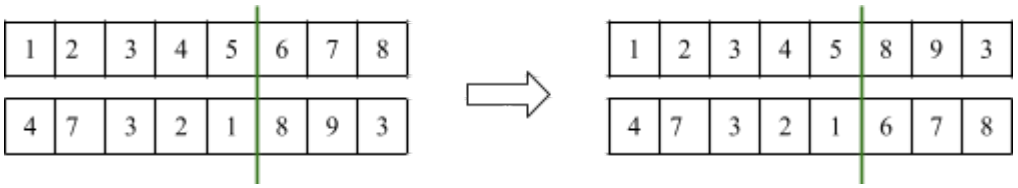


Figura 6. Representación del *crossover* de un único punto. La división se determina de forma aleatoria. Elaborado con Docs.google.com.

3.1.1 VERSIONES AG

Para conseguir el objetivo propuesto, se modificaron distintos parámetros del algoritmo, en especial la función de aptitud, resultando en las siguientes variantes.

AG 01

La primera versión del programa permite distinguir tan solo entre maligno, marcado con un 1, y sano o benigno, indicado con un 0. Se utiliza como función de aptitud la función sigmoide (Fig. 7), escogida porque varía entre 0 y 1, con posibilidad de deformarse. Se calculará de forma que

$$f(x_k) = \frac{1}{1 + e^{-g(x_k)}} \in (0, 1)$$

$$g(x_k) = \sum_{j=1}^{n} \gamma_j \bullet x_{kj}$$

siendo f la función sigmoide calculada para cada paciente al que pertenece un conjunto de muestras x_k, y $g(x_k)$ la función correspondiente al sumatorio desde 1 hasta el tamaño n de población seleccionado de la multiplicación de cada término $\gamma_j \in \mathbb{R}$ del individuo $\omega \in \mathbb{R}^7$ calculado por el AG por cada valor de biomarcador o atributo x_{kj} tomado del conjunto de datos. El individuo resultante sería de la forma

$$\omega = [\gamma_1, \gamma_2, \gamma_3, \gamma_4, \gamma_5, \gamma_6, \gamma_7].$$

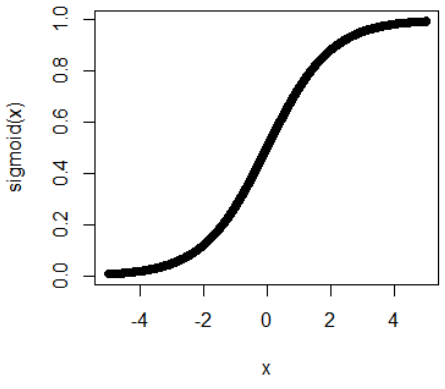


Figura 7. Función sigmoide. Dibujada utilizando RStudio.

Así se obtiene un número entre 0 y 1, que se resta a la salida esperada b_k para cada paciente k, calculando así una aptitud parcial $f_p(x_k)$ igual a la resta de 1 menos la diferencia absoluta. Cuanto menor sea la diferencia entre ambos valores, más ajustado estará el diagnóstico y mayor será la aptitud parcial. La aptitud total del individuo producido por el AG se calcula como el sumatorio desde 1 hasta el número m de pacientes de las aptitudes parciales correspondientes a la aplicación de las anteriores expresiones,

$$f_p(x_k) = 1 - \left| b_k - f(x_k) \right| \geq 0$$

$$F(\omega) = \sum_{k=1}^{m} f_p(x_k).$$

El código correspondiente a esta función de aptitud puede verse en la Figura 16 del Anexo I.

AG 02

La siguiente prueba mantiene el cálculo de la función sigmoide, variando tan solo la fórmula para la aptitud. En vez de calcular una aptitud parcial, la aptitud de cada individuo proviene de sumar 1 cada vez que la diferencia entre la salida esperada y la estimada sea menor al umbral u de forma que

$$F(\omega)_{t+1} = F(\omega)_t + 1 \; si \; \left| b_k - f(x_k) \right| \leq u \, .$$

Esto se traduce a código de la forma descrita en la Figura 17 del Anexo I.

AG 03

Para incrementar la utilidad del algoritmo se cambia la búsqueda de la capacidad de diferenciar entre maligno y sano o benigno a diferenciar entre los tres estados, siendo sano -1, benigno 0 y maligno 1. Se utilizan las fórmulas definidas para el AG 01, con la siguiente deformación de la función sigmoide para aumentar el rango de salidas

$$f(x_k) = \frac{2}{1 + e^{-g(x_k)}} - 1 \in (-1, 1).$$

AG 04

Con el fin de añadir variabilidad, se busca aumentar la pendiente, variando la población inicial y la función de mutación. La nueva población inicial incluye valores de r y α, del siguiente modo

$$f(x_k) = \frac{2}{1 + e^{-g(x_k)\bullet\alpha}} - 1 \in (-1, 1)$$

$$g(x_k) = \sum_{j=1}^{n} \gamma_j \bullet x_{kj}^{r_j},$$

siendo los intervalos de estos parámetros $\gamma_j \in [-1, 1]$, $r_j \in [1, 4]$ y $\alpha \in [1, 10]$.

Así, el individuo resultante sería de la forma

$$\omega = [\gamma_1, \gamma_2, \gamma_3, \gamma_4, \gamma_5, \gamma_6, \gamma_7, r_1, r_2, r_3, r_4, r_5, r_6, r_7, \alpha].$$

El cálculo de la mutación pasa a ser, para γ_j y α, el valor original multiplicado por 2, mientras que para los valores r_j se calcula como la mutación original. Se prueba también esta modificación en los algoritmos que dividen únicamente entre dos clases (AG 01), variando tan solo la función de la sigmoide, usando la original con salida en (0, 1).

AG 05

Buscando aumentar la capacidad de diagnóstico del algoritmo, se decide continuar haciendo los cálculos por poblaciones, de forma que se edita la base de datos para separarla en dos poblaciones, asignando en la Población 1 el 0 a los diagnósticos de tumor maligno y 1 al resto y, en la Población 2, 1 a los benignos y 0 a los demás, del modo descrito en la Figura 8.

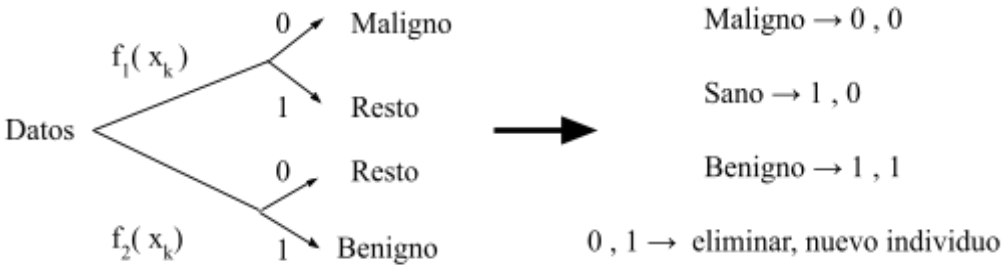


Figura 8. División de las bases de datos para el AG 05, elaborado con Docs.google.com.

La aptitud se calcula de forma conjunta, buscando qué dos individuos consiguen una mayor aptitud al ser aplicados ambos a los datos. Se aplica por separado la función sigmoide $f_1(x_k)$ y $f_2(x_k)$ para calcularlo como aptitudes parciales que después se suman en $f_{ap}(x_k)$, para obtener una salida estimada de entre 0 y 2. Para calcular la aptitud total se compara esta salida estimada con la esperada b_k, que ha sido modificada para ser 0 si el diagnóstico era de tumor maligno, 1 para ausencia del mismo y 2 para benigno. De forma matemática, esto sería

$$f_1(x_k) = \frac{1}{1 + e^{-g(x_k)\bullet\alpha}} \in (0, 1)$$

$$f_2(x_k) = \frac{1}{1 + e^{-g(x_k)\cdot\alpha}} \in (0, 1)$$

$$f_{ap}(x_k) = f_1(x_k) + f_2(x_k) \in (0, 2)$$

$$f_p(x_k) = 2 - \left| b_k - f_{ap}(x_k) \right| \geq 0$$

$$F(\omega) = \sum_{k=1}^{m} f_p(x_k)$$

Véase la Figura 18 en el Anexo I que muestra el código correspondiente a esta función de aptitud cuando se tiene un individuo con valores *r* y α descritos en el apartado anterior.

AG 06

Se modifica el algoritmo anterior dividido en poblaciones para obtener dos individuos con aptitudes calculadas de forma independiente, las aptitudes parciales calculadas no se suman, de forma que cada sigmoide está calculada entre 0 y 1 y se compara con las salidas codificadas como 0 y 1, obteniendo el mejor individuo, es decir, el de mayor aptitud, para separar entre maligno y resto con $f_1(x_k)$ y el mejor para separar entre benigno y el resto con $f_2(x_k)$. Solo se realiza el paso de suma a la hora de comprobar la eficacia con los datos de validación, realizando la misma comparación que en el AG 05.

3.2 REDES NEURONALES

Este algoritmo, inspirado por el comportamiento de las neuronas en el cerebro, fue originalmente creado para estudiar la función cerebral, y ha acabado por convertirse en uno de los más usados en ML, ignorando este propósito inicial (Greener *et al*., 2022). Basándose en el comportamiento de una neurona, que recibe distintas señales a través de las dendritas y envía un único potencial de acción por el axón, se idea un modelo en el que se combinan varios valores de entrada y se produce una única salida, lo que permite la clasificación (Kriegeskorte y Golan, 2019).

3.2.1 PERCEPTRÓN SIMPLE

Primero, se comenzará con una única neurona (Fig. 9), que sirve como discriminante lineal de los patrones de entrada, para la que se prueban las variaciones descritas a continuación. Se introducen una serie de datos (*s* muestras con *n* atributos), denominados *inputs* o entradas

$\underline{x_k} = (x_{k1}, \ldots, x_{kn}) \in \mathbb{R}^n$, que se combinan de forma linear con una serie de pesos $\underline{w} = (w_1, \ldots, w_n) \in \mathbb{R}^n$. La suma de estos pesos multiplicados se introduce en una función de activación no lineal que da un resultado entre 0 y 1, que será la salida esperada. El objetivo es que la salida esperada *d* sea lo más próxima posible a la salida obtenida *y*, calculándose el error entre ellas para cada muestra mediante el error cuadrático

$$E(k) = \frac{1}{2}(y_k - d_k)^2$$

(Silva *et al*., 2020), siendo el error cuadrático medio total

$$E = \frac{1}{s} \sum_{k=1}^{s} E(k).$$

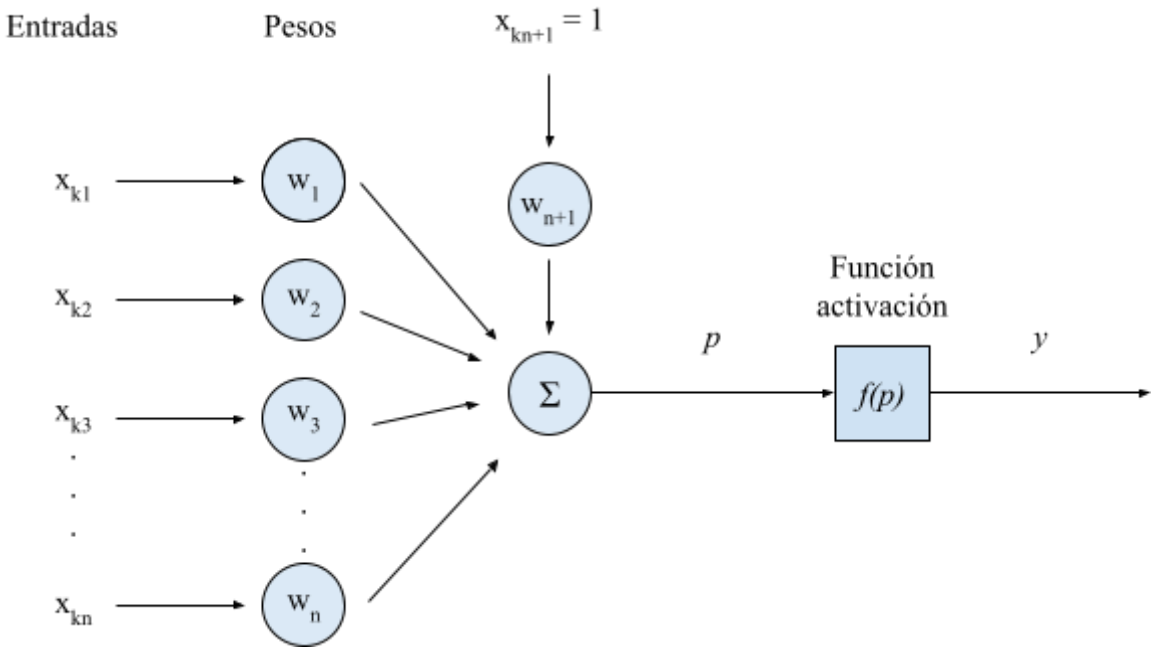


Figura 9. Representación del funcionamiento de una única neurona o perceptrón. En el presente estudio, cada *x* representaría un valor de un biomarcador, obteniéndose una *y* para cada conjunto de biomarcadores, es decir, para cada paciente. La n-upla (x_{k1},x_{k2},...,x_{kn}) corresponde al conjunto de atributos de cada paciente *k*. La conexión extra hace el papel de umbral. Para los datos del presente trabajo se tiene $n = 7$. Elaborado con Docs.google.com.

Partiendo de este concepto se escriben las siguientes versiones del algoritmo, buscando la clasificación de cada paciente en una de las tres clases.

Neurona 01

Se programan 6 opciones para la función de aproximación o activación, siendo estas

Opción 1 → $y = p \in (-\infty, \infty)$

Opción 2 → $y = \frac{2}{1+e^{-p}} \in (0,2)$

Opción 3 → $y = sin(p) + 1 \in (0,2)$

Opción 4 → $y = 2e^{-p^2} \in (0,2)$

Opción 5 → $y = \frac{2p}{1+p^2} + 1 \in (0,2)$

Opción 6 → $y = \frac{e^p - e^{-p}}{e^p + e^{-p}} + 1 \in (0,2)$

representando *y* la salida estimada, que se comparará con la salida esperada *d*, y siendo *p* el potencial calculado como

$$p = \sum_{j=1}^{n+1} w_j \bullet x_{kj} .$$

Mantenemos la notación x_{kj} como cada uno de los atributos contenidos en los datos y w_j serán los pesos. El cálculo del nuevo peso se realiza restando el inicial menos un *gamma (γ)*, determinado por el usuario, multiplicado por la diferencia entre la salida esperada y la obtenida, por la derivada de la función de aproximación y por la muestra, conocido como regla de aprendizaje del descenso de gradiente (Goodfellow *et al*., 2016)

$$w_j(t+1) = w_j(t) + (-\gamma \bullet (y-d) \bullet f'(p) \bullet x_{kj}).$$

Neurona 02

Se modifica la forma de tomar los datos para evitar parcialidad, colocando la base de datos para que pase por algoritmo en orden un conjunto de muestras de cada uno de los tres diagnósticos. Se cambia también el cálculo del potencial para evitar que sea lineal y por tanto más restrictivo, así como el cálculo del nuevo individuo, que pasa a ser

$$p = \sum_{j=1}^{n+1} w_j \bullet x_{kj}^{r_j}$$

$$w_j(t+1) = w_j(t) + (-\gamma \bullet (y-d) \bullet y'(p) \bullet x_{kj}^{r_j}).$$

Neurona 03

Se pretende en este algoritmo combinar ambos métodos, tomando el individuo inicial sobre el que comienza a actuar la neurona del mejor individuo de la historia del AG con los cálculos que mejor funcionan.

3.2.2 PERCEPTRÓN MULTICAPA

Para una mayor precisión, se añaden más neuronas a la red, creando un perceptrón multicapa (MLP) (Fig. 10). Esta serie de capas de neuronas que forma la red es capaz de aprender patrones complejos que se encuentran en un conjunto de datos dado. Las capas están interconectadas mediante estas neuronas, que transmiten información a través de conexiones a las que corresponden distintos pesos. Como sucede en el perceptrón simple, estos pesos van variando, y el algoritmo aprende aquellos con los que se obtiene un menor error, calculado con el error descrito en el apartado anterior, entre la salida esperada y la obtenida con el fin de obtener un modelo altamente predictivo (Alkadri *et al*., 2021).

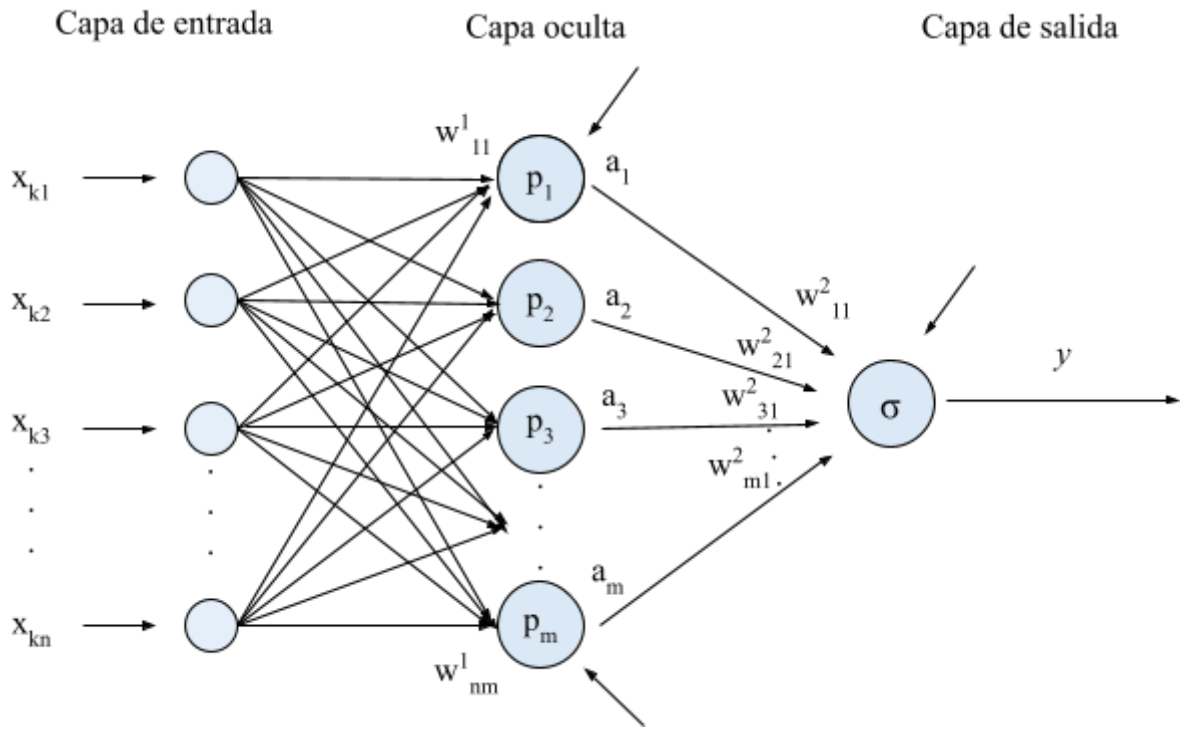


Figura 10. Representación de un MLP con una sola capa oculta. Los significados son iguales a los de la Figura 9, exceptuando σ, que en este caso representa la función sigmoide. En la capa 1, la función de activación usada también es la sigmoide. La notación w^c_{ij} representa el peso de la conexión de la neurona *i* a la neurona *i* de la capa *c* a la capa *c* + 1. Diagrama elaborado con Docs.google.com.

Para lograr este objetivo se usa el método del descenso de gradiente estocástico, pero esto requiere conocer el error. Para la función de aprendizaje de las capas ocultas de la red se utiliza la retropropagación, aplicando el gradiente de la última capa a la primera, ya que las capas ocultas no tienen un término de error que se pueda usar. Esto se hace aplicando la regla de la cadena, ya que las derivadas de los pesos de las capas ocultas están relacionadas con la predicción de los errores en la capa posterior. Esto lleva a la obtención de la Regla Delta Generalizada que define la modificación de los pesos (Goodfellow *et al*., 2016; Verguts, 2022). En la red usada en el presente trabajo se utilizará únicamente una capa oculta, cuya función de aprendizaje de modificación de los pesos w_{ij} para cada muestra *k* viene dada por

$$w_{ij}(t\ +\ 1)\ =w_{ij}(t)\ -\ \gamma\ \frac{\partial E(t)}{\partial w_{ij}}$$

siendo *t* el contador para el tiempo de entrenamiento y γ una tasa de aprendizaje determinada por el usuario.

3.3. REGRESIÓN LOGÍSTICA

El último tipo de algoritmo que se va a desarrollar para la comparación en este trabajo es la RL. Este método se usa para estimar la relación entre una o más variables independientes y una variable de salida binaria permitiendo, por tanto, estimar la probabilidad de un resultado en particular (Schober y Vetter, 2021). RL se lleva utilizando en diversos campos clínicos desde la década de los 70, más recientemente, por ejemplo, para identificar factores de riesgo de padecer COVID-19 (Trübner *et al*., 2021), predecir comorbilidades en menores con parálisis cerebral (Bertoncelli *et al*., 2020) o para clasificar tumores phyllodes de mama como malignos o benignos (Li *et al*., 2022).

En RL, para poder obtener estas probabilidades, las salidas binarias que introducimos como 0 (control) o 1 (maligno), se convierten en una función continua con imagen [0, 1] mediante la función logit, algebraicamente escrita como

$$logit(p_{x_k})\ =\ log(\frac{p_{x_k}}{1-p_{x_k}})\ =\ \beta_0\ +\ \beta_1 X_{k1}\ +\ \cdots\ +\ \beta_n X_{kn},$$

siendo β_j los parámetros estimados asociados a las variables introducidas X_{kj}, para los 7 atributos. Esto permite convertir una salida *Y* binaria en una probabilidad de la siguiente manera (Hosmer *et al*., 2013; Zabor *et al*., 2022)

$$P(Y\ =\ 1|X_k)\ =(\frac{exp(\beta_0+\beta_1 X_{k1}+\cdots+\beta_k X_{kn})}{1+exp(\beta_0+\beta_1 X_{k1}+\cdots+\beta_k X_{kn})}).$$

Se pretende que esta probabilidad estimada sea lo más cercana posible al diagnóstico esperado, indicando la precisión del modelo.

Una vez desarrollados los programas que utilizarán cada tipo de algoritmo, se procede a recoger los resultados obtenidos y compararlos entre sí y con el rendimiento de aquellos encontrados en la literatura.

4. RESULTADOS

Para comenzar, se estudian estadísticamente las distintas pruebas de los AG (Tabla 1) y aquellos que dividen entre dos clases se comparan también mediante la realización de curvas ROC (Fig. 11). Primero, se ha corrido cada uno de los programas probando distintos valores para las variables definidas por el usuario, con el fin de estimar cuáles se utilizarán para la comparación.

Los programas acompañados de "C/M" sirven para clasificar entre controles (C) y casos malignos (M), mientras que aquellos acompañados de "B/M" separan benignos (B) de malignos.

Para comparar, se dividen los datos en distintos subconjuntos. Se utilizarán distintos archivos para los algoritmos que dividan entre C/M, para B/M, y para las tres clases. Se han separado como se explicó previamente, resultando en un porcentaje del 33%, 30% y 40% de datos reservados para la predicción respectivamente para cada tipo de clasificación. Para cada algoritmo se están teniendo en cuenta 20 réplicas.

Se determinó que, en un AG, los parámetros más favorables son un tamaño de población de 75, un tiempo máximo de 100, una probabilidad de cruce y de mutación de 0.6 y 0.001 respectivamente, un valor z de 4 y un umbral de 0.499999.

Para cada AG se estudia la precisión, la sensibilidad y la especificidad del mismo en distintas pruebas para su posterior comparación. Para aquellos algoritmos que clasifican en tres clases, se indica la precisión, calculada como el número de diagnósticos correctos frente al total de la población a evaluar, especificidad, calculada como el número de verdaderos negativos frente a la suma de estos más falsos positivos, y sensibilidad, calculada como el número de verdaderos positivos frente a la suma de estos más falsos negativos.

Tabla 1. Valores medios de precisión para cada prueba, en porcentaje. Datos expresados para un intervalo de confianza (IC) del 95%.
a) Resultados para la clasificación entre C y M. **b)** Resultados para la clasificación entre B y M. **c)** Resultados para la clasificación entre C, B y M. Los AGs acompañados de ".2" indican que se usa el método descrito en el apartado correspondiente pero haciendo uso de los parámetros *a* y *r*, mientras que el resto contiene sólo lo explicado en su apartado. "Esp.": especificidad, "Sen.": sensibilidad, "C,B,M": control, benigno, maligno.

a)	AG01 C/M	AG02 C/M	AG04 C/M
Precisión	92.31 ± 2.31	87.75 ± 2.47	50.35 ± 2.94
Especificidad	86.66 ± 4.38	96.33 ± 9.30	43.00 ± 14.50
Sensibilidad	97.95 ± 0.93	79.18 ± 4.91	57.70 ± 14.40

b)	AG01 B/M	AG02 B/M	AG04 B/M
Precisión	70.52 ± 3.91	72.14 ± 3.24	48.80 ± 3.62
Especificidad	59.32 ± 11.70	69.09 ± 7.42	38.86 ± 11.80
Sensibilidad	81.73 ± 4.09	75.19 ± 3.40	58.75 ± 9.88

c)	AG03	AG03.2	AG05	AG05.2	AG06	AG06.2
Precisión	51.08 ± 2.45	37.97 ± 2.75	18.28 ± 1.31	25.22 ± 2.05	16.12 ± 0.48	65.95 ± 7.81
Esp. C	99.59 ± 0.56	96.12 ± 2.35	100.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00	95.19 ± 2.55
Sen. C	1.67 ± 1.56	5.56 ± 5.11	0.16 ± 0.30	0.39 ± 0.37	1.09 ± 0.87	28.44 ± 7.50
Esp. B	36.77 ± 4.88	22.74 ± 8.59	29.90 ± 11.20	69.95 ± 11.40	2.30 ± 1.95	33.37 ± 8.20
Sen. B	88.53 ± 4.18	86.32 ± 9.02	75.56 ± 8.99	25.83 ± 13.10	100.00 ± 0.00	85.28 ± 8.62
Esp. M	91.35 ± 3.05	90.77 ± 9.19	68.17 ± 11.70	30.12 ± 11.40	98.11 ± 2.37	87.87 ± 7.72
Sen. M	45.08 ± 6.32	21.41 ± 8.98	22.06 ± 8.81	71.62 ± 12.80	0.00 ± 0.00	13.82 ± 9.33

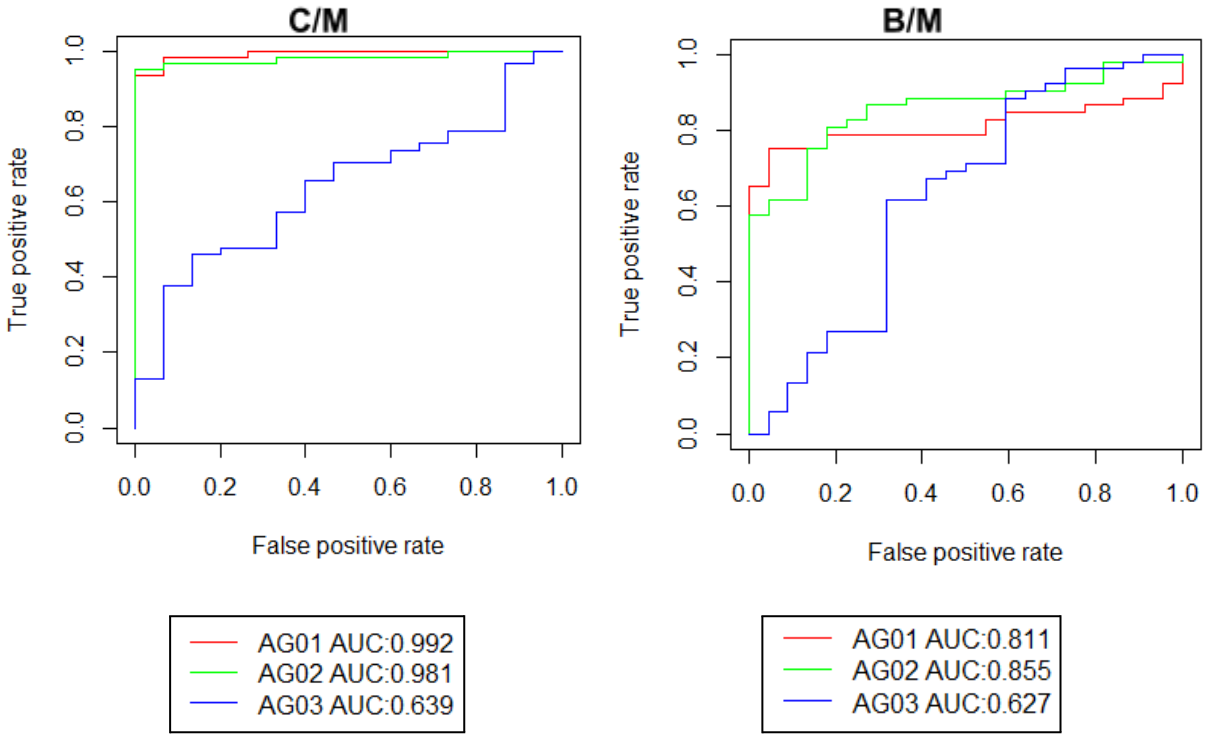


Figura 11. Curvas ROC para la clasificación entre C/M y B/M de los algoritmos AG01 (rojo), AG02 (verde) y AG04 (azul) y valor de AUC para cada uno. Representación de "*True positive rate*" o sensibilidad y "*False positive rate*" o 1 - especificidad.

A continuación, se procede a estudiar los resultados obtenidos con las redes neuronales. Para los perceptrones simples se selecciona un umbral de 0.2, un tiempo máximo de 1000 y un error absoluto mínimo de 0.01 para cada una de las opciones.

Para todas las variantes del perceptrón simple se obtiene que para todas las opciones de cálculo de la función de activación el porcentaje de aciertos se encuentra en torno al 61%, excepto la opción 5, descrita en la sección 3.2.1, para la que se obtiene tan solo alrededor de un 11%. Ese porcentaje se mantiene porque el algoritmo presenta parcialidad, identificando correctamente tan solo los individuos malignos. A pesar de añadir variabilidad, según lo indicado en la sección Neurona 02, tan solo puede destacarse la opción 2 que aumenta a alrededor del 68%. Al partir de los mejores individuos obtenidos del AG03 y del AG06, en la Neurona 03 existe algo más de variabilidad, dependiendo de los pesos iniciales indicados, pero no llega a superar el 60% de diagnósticos correctos. Además, las distintas réplicas para un mismo peso inicial arrojan el mismo resultado, a pesar de aumentar el tiempo máximo.

Por otro lado, los resultados del MLP a la hora de separar entre las tres clases pueden verse en la Tabla 2a mientras que los resultados de separar C/M y B/M están descritos en la Tabla 2b.

Tabla 2a. Medias de las 20 réplicas del MLP para la separación en tres clases, con 8 neuronas en la capa oculta, una tasa de aprendizaje de 0.5 y un tiempo de 100 por el número de muestras en entrenamiento. Valores expresados en porcentaje excepto el error medio. Datos expresados para un IC del 95%.

MLP			
Error medio	0.39 ± 0.10	**Esp. B**	98.41 ± 0.972
Precisión	39.70 ± 9.64	**Sen. B**	1.91 ± 1.59
Esp. C	40.20 ± 16.10	**Esp. M**	80.67 ± 7.94
Sen. C	88.61 ± 5.13	**Sen. M**	46.02 ± 18.40

Tabla 2b. Medias de las 20 réplicas del MLP para la separación de C/M y B/M, expresadas en porcentaje exceptuando el error medio, con 8 neuronas en la capa oculta, una tasa de aprendizaje de 0.5 y un tiempo de 100 por el número de muestras en entrenamiento. Datos expresados para un IC del 95%

	MLP C/M	**MLP B/M**
Error medio	0.09 ± 0.06	0.10 ± 0.03
Precisión	89.39 ± 8.31	76.89 ± 4.86
Especificidad	90.09 ± 10.30	77.05 ± 4.65
Sensibilidad	88.69 ± 6.80	76.73 ± 6.24

Por último, se estudian también los resultados de las réplicas de regresión logística, que pueden ser observados en la Tabla 3.

Tabla 3. Resultados de la regresión lineal, utilizando los mismos archivos que para los AG y MLP C/M y B/M, en porcentaje. Se obtiene un único resultado.

	RL C/M	**RL B/M**
Precisión	84.21	72.97
Especificidad	80.33	73.08
Sensibilidad	100	72.73

Para finalizar, se realiza una comparación de los mejores métodos de cada tipo de algoritmo mediante curvas ROC (Fig. 12, Fig. 13).

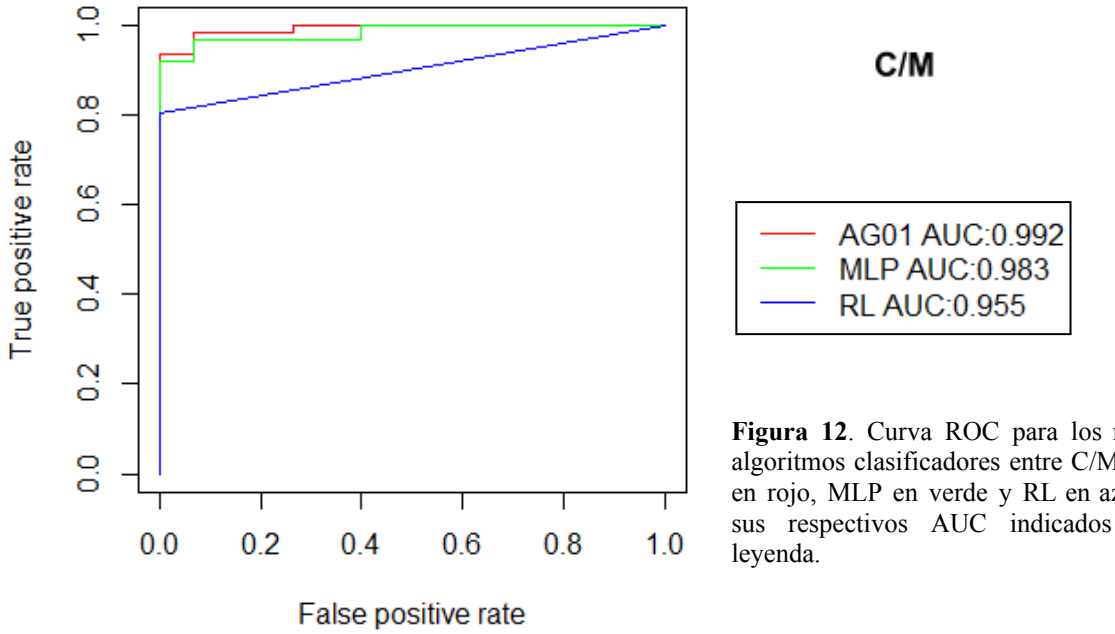


Figura 12. Curva ROC para los mejores algoritmos clasificadores entre C/M, AG01 en rojo, MLP en verde y RL en azul, con sus respectivos AUC indicados en la leyenda.

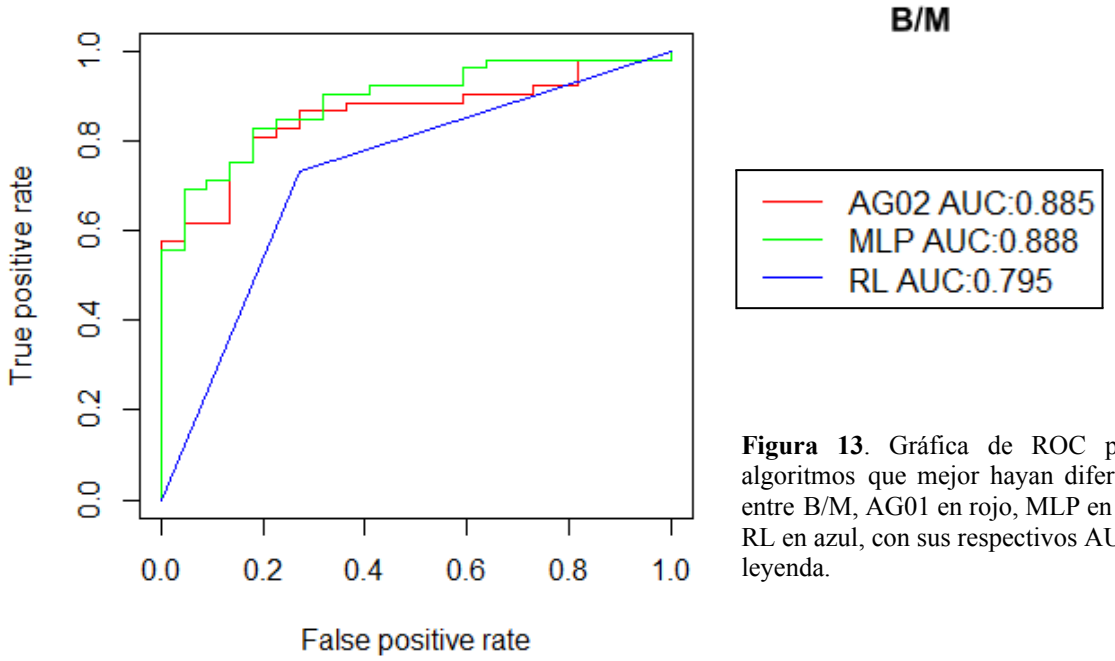


Figura 13. Gráfica de ROC para los algoritmos que mejor hayan diferenciado entre B/M, AG01 en rojo, MLP en verde y RL en azul, con sus respectivos AUC en la leyenda.

Una vez determinados los mejores modelos, se procede a realizar una comparación estadística de los mismos, utilizando el test de McNemar (McNemar, 1947). La hipótesis nula de partida es que el AG no tiene más capacidad predictiva que el RL o MLP, por lo que enfrentaremos a ambos con este primero, comparando también los otros valores disponibles para la clasificación entre C/M (Tabla 4a) y B/M (Tabla 4b).

Tabla 4a. Especificidad, sensibilidad y valores *p* resultantes de aplicar el test de McNemar a las parejas de algoritmos indicadas para cada tipo de algoritmo, siendo AG AG01, separando C/M..

Método	**AG**	**MLP**	**RL**
Especificidad (95% IC)	0.86 (0.82-0.91)	0.9 (0.79-1)	0.8
Sensibilidad (95% IC)	0.98 (0.97-0.99)	0.89 (0.81-0.95)	1
Valor *p*		0.06543	0.01294

Tabla 4b. Especificidad, sensibilidad y valores *p* resultantes de aplicar el test de McNemar a las parejas de algoritmos indicadas para cada tipo de algoritmo, siendo AG AG02, separando B/M.

Método	**AG**	**MLP**	**RL**
Especificidad (95% CI)	0.69 (0.62-0.77)	0.77 (0.72-0.82)	0.73
Sensibilidad (95% CI)	0.75 (0.72-0.79)	0.77 (0.7-0.83)	0.73
Valor *p*		0.62906	0.35928

Si el valor de $p > 0.05$ se trata de un valor no significativo, por lo que se puede aceptar la hipótesis nula. Esto sucede al comparar la capacidad predictiva del AG y del MLP, y la del AG y el RL en cuanto a la clasificación de B y M.

Por el contrario, si el valor $p < 0.05$ se rechaza la hipótesis nula, pudiendo afirmar que en cuanto a la separación de casos control y casos con tumor maligno es mejor utilizar un AG frente a un RL.

Por todo esto, no se puede determinar que la aplicación del AG resulte en una predicción mucho mejor a la que se puede obtener con otros algoritmos, a pesar de que resulte en la mayor precisión a la hora de clasificar en C/M (Tabla 1a). Además, este método presenta otras desventajas, como por ejemplo más tiempo de procesamiento a la hora de calcular el mejor individuo.

Para la selección de un algoritmo clasificador se debe tener en cuenta el buscar una mayor sensibilidad o una mayor especificidad dependiendo entre otras cosas de la enfermedad para la que se está realizando el diagnóstico. Para aquellas que son frecuentes, es más adecuado un método con mayor sensibilidad para conseguir el menor número posible de falsos negativos.

Al contrario, para PDAC sería más apropiado usar aquel método que tenga mayor especificidad, resultando en la menor cantidad posible de falsos positivos, ya que la confirmación de la enfermedad se trata de un proceso bastante invasivo y probablemente

costoso. Como puede comprobarse, resulta más sencillo clasificar entre casos control y aquellos que tienen un tumor maligno, lo que puede resultar útil si se pretende usar este tipo de prueba en la población general. Aún así, podría ser de más interés ser capaces de diferenciar entre aquellos tumores que sean benignos y malignos, pudiendo con tan solo una muestra de orina y un análisis de sangre evaluar cada cierto tiempo el progreso de la enfermedad.

Sería ideal un algoritmo que permitiera distinguir entre los tres estados, como los descritos en las Tablas 1c y 2a, siendo el que más se aproxima a las características buscadas el AG06.2, con una precisión del 65.95%, una especificidad del 95.19% para casos control y del 87.87% para tumores malignos, pero están acompañadas de unas sensibilidades muy bajas, por lo que no compensa su uso, llevando a una sobre-estimación exagerada del riesgo. Por esto, parecen más indicados aquellos que llevan a cabo una clasificación binaria, pudiendo aplicar uno u otro según el estado del paciente.

Por último, se pueden comparar los resultados obtenidos con los descrito en Debernardi y colaboradores (2020), artículo para el que se usaron los mismos datos. En la Figura 14a puede verse la curva ROC realizada al pasar los datos por su algoritmo PancRISK, que utiliza la RL, siendo de interés concretamente la correspondiente al panel junto con CA19-9, ya que son los datos usados en este trabajo. Esta gráfica se corresponde a la clasificación entre control y PDAC. El AUC es de 0.993, prácticamente igual al que se obtiene usando un AG, de 0.992, como puede verse en la Figura 12. La sensibilidad también es similar, siendo la del artículo de 0.971, mientras que la especificidad es mayor que en este estudio, 0.978 comparado con 0.866 (Tabla 1a).

En la Figura 14b puede verse la curva correspondiente a la clasificación entre benigno y PDAC. El AUC para la misma es de 0.919, comparado con 0.885 obtenido por el AG, que consigue además menor especificidad y sensibilidad que las obtenidas en el artículo.

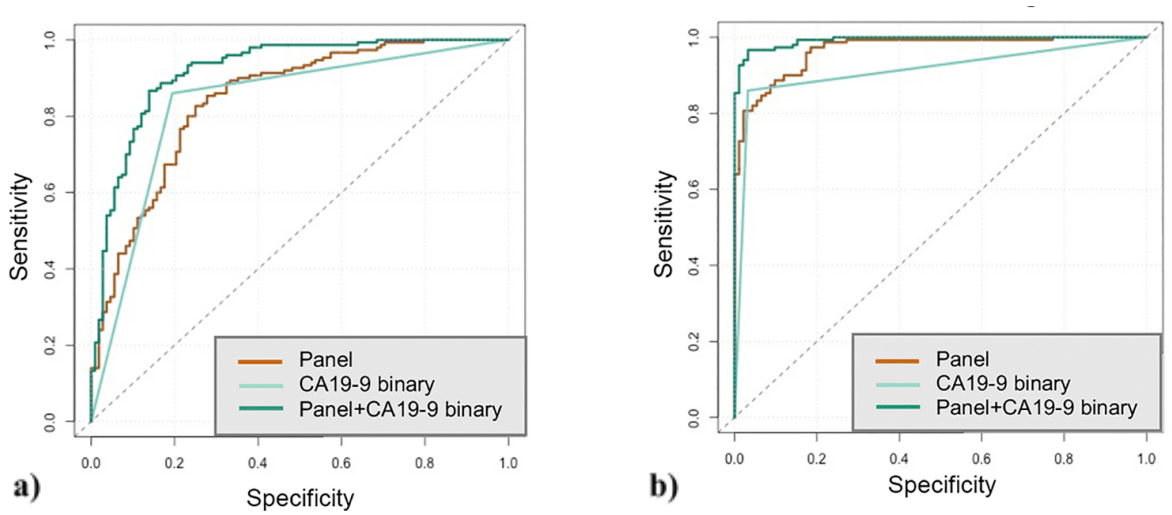


Figura 14. Gráficas ROC para la clasificación entre casos **a)** control y con PDAC y **b)** benignos y con PDAC, utilizando el algoritmo PancRISK para el panel de biomarcadores descrito en la introducción, CA19-9 o ambos. Gráficas tomadas de Debernardi *et al*., 2020.

Por esto, podemos afirmar que, en general, PancRISK tiene mejor capacidad de predicción que el algoritmo desarrollado en este trabajo, aunque los resultados sean comparables en algunas áreas.

5. CONCLUSIONES Y VÍAS FUTURAS

La capacidad de detectar el PDAC en un estadío temprano es crucial para el pronóstico de los pacientes, siendo extremadamente útil el uso de técnicas de diagnóstico poco invasivas, pudiendo destacar aquellas basadas en biomarcadores como las más prometedoras.

Será de interés continuar estudiando otros marcadores que puedan llevar a un diagnóstico aún más preciso, pudiendo considerar los estudiados recientemente por Hrabák y colaboradores (2022), de los que se pueden destacar S100A11 y DJ-1. También se podrían probar modificaciones de los algoritmos, entre otras, el uso de otros operadores de los descritos en la Figura 3 para el AG o la adición de más capas ocultas en el MLP.

Incluso, podrían utilizarse otras técnicas de *machine learning*, siendo las SVM las consideradas más precisas al tratarse del PDAC. Actualmente, aunque el estudio de biomarcadores presente la ventaja de una menor invasividad, la mayor parte de aplicaciones en la detección de cáncer se centran en el uso de estas técnicas para la clasificación desde imágenes, utilizando métodos más novedosos y con cálculos más complejos, que requieren ordenadores más potentes, como los autocodificadores variacionales o las redes neuronales complejas (Painuli *et al*., 2022; Liu *et al*., 2022).

Una de las claves del avance, no solo de algoritmos de detección de PDAC, si no de cualquier campo de la bioinformática, se podría considerar también la puesta en común de los datos, permitiendo que los algoritmos entrenen con una mayor cantidad de muestras, siendo estas además más variados y pudiendo evitar parcialidad causada por las pruebas usadas en un determinado hospital o zona a la que pertenezcan los pacientes estudiados.

A pesar de las altas expectativas puestas sobre este tipo de herramienta, aún existen ciertas limitaciones antes de poder proceder a su aplicación clínica, teniendo en cuenta cuestiones éticas y legales como sobre quién recae la responsabilidad del diagnóstico en caso de errores y cuánto deben saber los profesionales clínicos y los pacientes sobre los algoritmos usados, así como cuestiones de protección de datos (Ngiam y Khor, 2019).

En definitiva, según los resultados obtenidos, hay varias opciones de algoritmos que permitirán poder usar IA y la aproximación de datos para conseguir un diagnóstico a partir los biomarcadores estudiados, creando una herramienta que podrá utilizarse para el cribado de la población de riesgo, estando un paso más cerca de reducir la mortalidad causada por el PDAC.

6. REFERENCIAS

Adeoye, J., Tan, J. Y., Choi, S. W. y Thomson, P. (2021) "Prediction models applying machine learning to oral cavity cancer outcomes: A systematic review", *International journal of medical informatics*, 154:104557. doi:10.1016/J.IJMEDINF.2021.104557.

Alkadri, S., Ledwos, N., Mirchi, N., Reich, A. *et al.* (2021) "Utilizing a multilayer perceptron artificial neural network to assess a virtual reality surgical procedure", *Computers in Biology and Medicine*, 136, p. 104770. doi:10.1016/J.COMPBIOMED.2021.104770.

Bertoncelli, C. M., Altamura, P., Vieira, E. R., Iyengar, S. S. *et al.* (2020) "PredictMed: A logistic regression–based model to predict health conditions in cerebral palsy", *Health Informatics Journal*, 26(3), pp. 2105–2118. doi:10.1177/1460458219898568.

Blyuss, O., Zaikin, A., Cherepanova, V., Munblit, D. *et al.* (2020) "Development of PancRISK, a urine biomarker-based risk score for stratified screening of pancreatic cancer patients", *British Journal of Cancer*, 122(5), pp. 692–696. doi:10.1038/s41416-019-0694-0.

Brezgyte, G., Shah, V., Jach, D. y Crnogorac‐jurcevic, T. (2021) "Non-Invasive Biomarkers for Earlier Detection of Pancreatic Cancer-A Comprehensive Review", *Cancers*, 13(11). doi:10.3390/CANCERS13112722.

Del Chiaro, M., Verbeke, C., Salvia, R., Klöppel, G. *et al.* (2013) "European experts consensus statement on cystic tumours of the pancreas", *Digestive and Liver Disease*, 45(9), pp. 703–711. doi:10.1016/J.DLD.2013.01.010.

Cohen, J. D., Li, L., Wang, Y., Thoburn, C. *et al*. (2018) "Detection and localization of surgically resectable cancers with a multi-analyte blood test", *Science (New York, N.Y.)*. Science, 359(6378), pp. 926–930. doi:10.1126/SCIENCE.AAR3247.

Collisson, E. A., Bailey, P., Chang, D. K. y Biankin, A. V. (2019) "Molecular subtypes of pancreatic cancer", *Nature Reviews Gastroenterology and Hepatology*, 16(4), pp. 207–220. doi:10.1038/S41575-019-0109-Y.

Davis, J. (2020) *Kaggle - Urinary biomarkers for pancreatic cancer*. Disponible en: https://www.kaggle.com/datasets/johnjdavisiv/urinary-biomarkers-for-pancreatic-cancer. (Accedido: 10 de 10 de 2021).

Debernardi, S., O'Brien, H., Algahmdi, A. S., Malats, N. *et al*. (2020) "A combination of urinary biomarker panel and PancRISK score for earlier detection of pancreatic cancer: A case-control study", *PLoS medicine*, 17(12). doi:10.1371/JOURNAL.PMED.1003489.

Docs.google.com (2022) *Google Docs.* Disponible en https://docs.google.com/ (Accedido: 10 de 09 de 2021).

Faradonbeh, R. S., Hasanipanah, M., Amnieh, H. B., Armaghani, D. J. *et al.* (2018) "Development of GP and GEP models to estimate an environmental issue induced by blasting operation", *Environmental Monitoring and Assessment*, 190(6). doi:10.1007/S10661-018-6719-Y.

Ghaheri, A., Shoar, S., Naderan, M. y Hoseini, S. S. (2015) "The Applications of Genetic Algorithms in Medicine", *Oman Medical Journal*, 30(6), p. 406. doi:10.5001/OMJ.2015.82.

Ghosh, M., Adhikary, S., Ghosh, K. K., Sardar, A. *et al*. (2019) "Genetic algorithm based cancerous gene identification from microarray data using ensemble of filter methods", *Medical & biological engineering & computing*, 57(1), pp. 159–176. doi:10.1007/S11517-018-1874-4.

Goggins, M., Overbeek, K. A., Brand, R., Syngal, S. *et al*. (2020) "Management of patients with increased risk for familial pancreatic cancer: updated recommendations from the International Cancer of the Pancreas Screening (CAPS) Consortium", *Gut*, 69(1), pp. 7–17. doi:10.1136/GUTJNL-2019-319352.

Goodfellow, I., Bengio, Y. y Courville, A. (2016) *Deep learning*. Cambridge: MIT Press.

Greener, J. G., Kandathil, S. M., Moffat, L. y Jones, D. T. (2022) "A guide to machine learning for biologists", *Nature Reviews Molecular Cell Biology*, 23(1), pp. 40–55. doi:10.1038/S41580-021-00407-0.

Harris, C.R., Millman, K.J., van der Walt, S.J., Gommers, R. *et al*. (2020) "Array programming with NumPy" *Nature,* 585, pp. 357–362. doi: 10.1038/s41586-020-2649-2.

Hrabák, P., Šoupal, J., Kalousová, M., Krechler, T. *et al*. (2022) "Novel biochemical markers for non-invasive detection of pancreatic cancer", *Neoplasma*, 69(2), pp. 474–483. doi: 10.4149/neo_2022_210730N1075

Hosmer, D. W., Lemeshow, S. y Sturdivant, R. X. (2013) *Applied Logistic Regression*. 3.ª ed. Hoboken:Wiley.

Hu, J. X., Helleberg, M., Jensen, A. B., Brunak, S. *et al*. (2019) "A Large-Cohort, Longitudinal Study Determines Precancer Disease Routes across Different Cancer Types", *Cancer research*, 79(4), pp. 864–872. doi:10.1158/0008-5472.CAN-18-1677.

Jackson, D. G. (2003) "The Lymphatics Revisited: New Perspectives from the Hyaluronan Receptor LYVE-1", *Trends in Cardiovascular Medicine*, 13(1), pp. 1–7. doi:10.1016/S1050-1738(02)00189-5.

Katoch, S., Chauhan, S. S. y Kumar, V. (2021) "A review on genetic algorithm: past, present, and future", *Multimedia tools and applications*, 80(5), pp. 8091–8126. doi:10.1007/S11042-020-10139-6.

Kenner, B., Chari, S. T., Kelsen, D., Klimstra, D. S. *et al*. (2021) "Artificial Intelligence and Early Detection of Pancreatic Cancer: 2020 Summative Review", *Pancreas*, 50(3), p. 251. doi:10.1097/MPA.0000000000001762.

Klett, H., Fuellgraf, H., Levit-Zerdoun, E., Hussung, S. *et al*. (2018) "Identification and validation of a diagnostic and prognostic multi-gene biomarker panel for pancreatic ductal adenocarcinoma", *Frontiers in Genetics*, 9(APR). doi:10.3389/FGENE.2018.00108/PDF.

Kriegeskorte, N. y Golan, T. (2019) "Neural network models and deep learning", *Current Biology*, 29(7), pp. R231–R236. doi:10.1016/J.CUB.2019.02.034.

Kruskal, W. H. y Wallis, W. A. (1952) "Use of Ranks in One-Criterion Variance Analysis", *Journal of the American Statistical Association*, 47(260), pp. 583-621.

Li, P., Cong, Z., Qiang, Y., Xiong, L. *et al.* (2018) "Clinical significance of CCBE1 expression in lung cancer", *Molecular Medicine Reports*, 17(2), pp. 2107–2112. doi:10.3892/MMR.2017.8187.

Li, Q., Wang, H., Zogopoulos, G., Shao, Q. *et al.* (2016) "Reg proteins promote acinar-to-ductal metaplasia and act as novel diagnostic and prognostic markers in pancreatic ductal adenocarcinoma", *Oncotarget*, 7(47), pp. 77838–77853. doi:10.18632/ONCOTARGET.12834.

Li, T., Li, Y., Yang, Y., Li, J. *et al*. (2022) "Logistic regression analysis of ultrasound findings in predicting the malignant and benign phyllodes tumor of breast", *PloS one*, 17(3). doi:10.1371/JOURNAL.PONE.0265952.

Liu, Y., Li, S., y Liu, Y. (2022) "Machine Learning-Driven Multiobjective Optimization: An Opportunity of Microfluidic Platforms Applied in Cancer Research", *Cells*, 11(5), 905. https://doi.org/10.3390/cells11050905

Loveday, B. P. T., Lipton, L. y Thomson, B. N. (2019) "Pancreatic cancer: An update on diagnosis and management", *Australian journal of general practice*, 48(12), pp. 826–831. doi:10.31128/AJGP-06-19-4957.

Luo, G., Jin, K., Deng, S., Cheng, H. *et al*. (2021) "Roles of CA19-9 in pancreatic cancer: Biomarker, predictor and promoter", *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer*, 1875(2), p. 188409. doi:10.1016/J.BBCAN.2020.188409.

Lyell, D., Coiera, E., Chen, J., Shah, P. *et al*. (2021) "How machine learning is embedded to support clinician decision making: An analysis of FDA-approved medical devices", *BMJ Health and Care Informatics*, 28(1). doi:10.1136/BMJHCI-2020-100301.

Lynch, C. M., Abdollahi, B., Fuqua, J. D., de Carlo, A. R. *et al*. (2017) "Prediction of lung cancer patient survival via supervised machine learning classification techniques", *International journal of medical informatics*, 108, pp. 1–8. doi:10.1016/J.IJMEDINF.2017.09.013.

McGuigan, A., Kelly, P., Turkington, R. C., Jones, C. *et al*. (2018) "Pancreatic cancer: A review of clinical diagnosis, epidemiology, treatment and outcomes", *World Journal of Gastroenterology*, 24(43), pp. 4846–4861. doi:10.3748/WJG.V24.I43.4846.

McNemar, Q. (1947) "Note on the sampling error of the difference between correlated proportions or percentages", *Psychometrika*, 12, pp. 153–157. doi:10.1007/BF02295996.

Ngiam, K. Y. y Khor, I. W. (2019) "Big data and machine learning algorithms for health-care delivery", *The Lancet Oncology*, 20(5), pp. e262–e273. doi:10.1016/S1470-2045(19)30149-4.

O'Neill, R. S. y Stoita, A. (2021) "Biomarkers in the diagnosis of pancreatic cancer: Are we closer to finding the golden ticket?", *World journal of gastroenterology*, 27(26), pp. 4045–4087. doi:10.3748/WJG.V27.I26.4045.

Painuli, D., Bhardwaj, S., y Köse, U. (2022) "Recent advancement in cancer diagnosis using machine learning and deep learning techniques: A comprehensive review", *Computers in biology and medicine*, 146, 105580. https://doi.org/10.1016/j.compbiomed.2022.105580.

Pereira, S. P., Oldfield, L., Ney, A., Hart, P. A. *et al.* (2020) "Early detection of pancreatic cancer", *The Lancet Gastroenterology & Hepatology*, 5(7), pp. 698–710. doi:10.1016/S2468-1253(19)30416-9.

Poruk, K. E., Gay, D. Z., Brown, K., Mulvihill, J. D. *et al*. (2013) "The Clinical Utility of CA 19-9 in Pancreatic Adenocarcinoma: Diagnostic and Prognostic Updates", *Current molecular medicine*, 13(3), p. 340. doi:10.2174/1566524011313030003.

Radon, T. P., Massat, N. J., Jones, R., Alrawashdeh, W. *et al*. (2015) "Identification of a Three-Biomarker Panel in Urine for Early Detection of Pancreatic Adenocarcinoma", *Clinical cancer research : an official journal of the American Association for Cancer Research*, 21(15), pp. 3512–3521. doi:10.1158/1078-0432.CCR-14-2467.

Reback, J., McKinney, W., Van den Bossche, J., Augspurger, T. *et al*. (2022) "pandas-dev/pandas: Pandas 1.4.1", *Zenodo*. doi:10.5281/zenodo.6053272.

RStudio Team (2022) *RStudio: Integrated Development Environment for R (Versión 2021.09.2+382) [Programa de ordenador]*. Disponible en: https://www.rstudio.com/products/rstudio/download/ (Accedido: 17 de 09 de 2021).

Schober, P. y Vetter, T. R. (2021) "Logistic Regression in Medical Research", *Anesthesia and Analgesia*, 132(2), p. 365. doi:10.1213/ANE.0000000000005247.

Siegel, R. L., Miller, K. D., Fuchs, H. E. y Jemal, A. (2022) "Cancer statistics, 2022", *CA: A Cancer Journal for Clinicians*, 72(1), pp. 7–33. doi:10.3322/CAAC.21708.

Silva, F., Sanz, M., Seixas, J., Solano, E. *et al*. (2020) "Perceptrons from memristors", *Neural networks : the official journal of the International Neural Network Society*, 122, pp. 273–278. doi:10.1016/J.NEUNET.2019.10.013.

Sing, T., Sander, O., Beerenwinkel, N. y Lengauer, T. (2005) "ROCR: visualizing classifier performance in R.", *Bioinformatics*, 21(20), 7881. doi:10.1093/BIOINFORMATICS/BTI623.

Stott, M. C., Oldfield, L., Hale, J., Costello, E. *et al.* (2022) "Recent advances in understanding pancreatic cancer", *Faculty reviews*, 11, p. 9. doi:10.12703/R/11-9.

Tang, M., Gao, L., He, B. y Yang, Y. (2022) "Machine Learning-Based Prognostic Prediction Models of Non-Metastatic Colon Cancer: Analyses Based on Surveillance, Epidemiology and End Results Database and a Chinese Cohort", *Cancer management and research*, 14, pp. 25–35. doi:10.2147/CMAR.S340739.

The Dia Developers (2014) *Dia (Versión 0.97.2) [Programa de ordenador]*. Disponible en: http://dia-installer.de/ (Accedido: 11 de 04 de 2022).

Trübner, F., Steigert, L., Echterdiek, F., Jung, N. *et al.* (2021) "Predictors of COVID-19 in an outpatient fever clinic", *PLoS ONE*, 16(July). doi:10.1371/JOURNAL.PONE.0254990.

Tseng, H. H., Wei, L., Cui, S., Luo, Y. *et al*. (2020) "Machine Learning and Imaging Informatics in Oncology", *Oncology (Switzerland)*, 98(6), pp. 344–362. doi:10.1159/000493575.

Ushey, K., Allaire, J. y Tang, Y. (2022) *reticulate: Interface to 'Python' (Versión 1.25) [Programa de ordenador]*. Disponible en: https://rstudio.github.io/reticulate/ (Accedido 17 de 09 de 2021).

Van Rossum, G. (2020) *Python: math — Mathematical functions (Versión 3.8.2) [Programa de ordenador]*. Fredericksburg: Python Software Foundation.

Verguts, T. (2022) *Introduction to Modeling Cognitive Processes*. Cambridge: MIT Press.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M. *et al*. (2020) "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python", *Nature Methods*, 17(3), 261-272. doi:10.1038/s41592-019-0686-2

Wada, K., Takaori, K. y Traverso, L. W. (2015) "Screening for Pancreatic Cancer", *Surgical Clinics of North America*, 95(5), pp. 1041–1052. doi:10.1016/J.SUC.2015.05.010.

Wang, Y., Zhou, Y., Tang, P., Shen, W., Fishman, E.K., Yuille, A.L. (2018) "Training Multi-organ Segmentation Networks with Sample Selection by Relaxed Upper Confident Bound" en Frangi, A., Schnabel, J., Davatzikos, C., Alberola-López, C., Fichtinger, G. (eds) *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*. Cham:Springer, pp 434–442.

Whitley, D. (2019) "Next Generation Genetic Algorithms: A User's Guide and Tutorial" en Gendreau, M. y Potvin, J. (eds.) *Handbook of Metaheuristics*. 3.ª ed. Cham: Springer, pp. 245-274.

Wiest, N. E., Moktan, V. P., Oman, S. P. y Chirilă, R. M. (2020) "Screening for pancreatic cancer: a review for general clinicians", *ROM. J. INTERN. MED*, 58, pp. 119–128. doi:10.2478/rjim-2020-0009.

Yamaguchi, J., Yokoyama, Y., Kokuryo, T., Ebata, T. *et al*. (2018) "Trefoil factor 1 inhibits epithelial-mesenchymal transition of pancreatic intraepithelial neoplasm", *Journal of Clinical Investigation*, 128(8), pp. 3619–3629. doi:10.1172/JCI97755.

Yang, J., Xu, R., Wang, C., Qiu, J. *et al*. (2021) "Early screening and diagnosis strategies of pancreatic cancer: a comprehensive review", *Cancer Communications*, 41(12), pp. 1257–1274. doi:10.1002/CAC2.12204.

Zabor, E. C., Reddy, C. A., Tendulkar, R. D. y Patil, S. (2022) "Logistic Regression in Clinical Studies", *International Journal of Radiation Oncology, Biology, Physics*, 112(2), pp. 271–277. doi:10.1016/J.IJROBP.2021.08.007.

Zhang, L., Sanagapalli, S. y Stoita, A. (2018) "Challenges in diagnosis of pancreatic cancer", *World journal of gastroenterology*, 24(19), pp. 2047–2060. doi:10.3748/WJG.V24.I19.2047.