

**UNIVERSIDAD COMPLUTENSE DE MADRID**  
**FACULTAD DE MEDICINA**



**TESIS DOCTORAL**

**Medicina de precisión en hematología: Inteligencia Artificial  
para el diagnóstico diferencial de anemias**

MEMORIA PARA OPTAR AL GRADO DE DOCTOR

PRESENTADA POR

**Roberto Oswaldo Trelles Martínez**

Directores

**Fernando Ataulfo González Fernández**  
**Nicolás Guil Mata**

Madrid

© Roberto Oswaldo Trelles Martínez, 2024

**UNIVERSIDAD COMPLUTENSE DE MADRID**

FACULTAD DE MEDICINA



**TESIS DOCTORAL**

**Medicina de precisión en hematología: Inteligencia Artificial para el diagnóstico**

**diferencial de anemias**

MEMORIA PARA OPTAR AL GRADO DE DOCTOR

**PRESENTADA POR:**

ROBERTO OSWALDO TRELLES MARTINEZ

**DIRECTORES:**

FERNANDO ATAULFO GONZÁLEZ FERNÁNDEZ

NICOLÁS GUIL MATA

**UNIVERSIDAD COMPLUTENSE DE MADRID**

**FACULTAD DE MEDICINA**



**TESIS DOCTORAL**

**Medicina de precisión en hematología: Inteligencia Artificial para el diagnóstico**

**diferencial de anemias**

**MEMORIA PARA OPTAR AL GRADO DE DOCTOR**

**PRESENTADA POR:**

**ROBERTO OSWALDO TRELLES MARTINEZ**

**DIRECTORES:**

**FERNANDO ATAULFO GONZÁLEZ FERNÁNDEZ**

**NICOLÁS GUIL MATA**

**PROGRAMA DE DOCTORADO EN INVESTIGACIÓN MEDICO-QUIRURGICA.**

**Hospital Universitario Clínico San Carlos de Madrid (HCSC)**

## **AGRADECIMIENTOS**

A María, mi compañera de vida, por su insuperable apoyo incondicional y por ser simplemente ella.

A mi madre Maite, por confiar siempre en mí, incluso cuando me he tambaleado.

A mi hermano Mario, por su gran ayuda con la organización y su mindfulness que siempre me empuja a sacar mi versión más ganadora.

A mis directores y a Sergio por su trabajo en equipo.

A Eulalia, la bibliotecaria, por ofrecerme su ayuda sin pedirme nada más que compromiso. Ella ha sido una gran inspiración.

A mi padre, Oswaldo Trelles Salazar, quien plantó la semilla de esta tesis doctoral, participó activamente y me ha acompañado durante todo el proceso. Una gran parte de esta tesis es tuya.

## INDICE

<b>1. Resumen .....</b>	<b>14</b>
<b>2. Abstract.....</b>	<b>19</b>
<b>3. Introducción .....</b>	<b>23</b>
3.1 Introducción a las anemias.....	28
3.1.1 Concepto, importancia y prevalencia .....	28
3.1.2 Bases del diagnóstico actual de las anemias .....	30
3.1.2.1 El hemograma.....	30
3.1.2.2. El frotis de sangre periférica .....	30
3.1.2.3 Electroforesis capilar y cromatografía líquida de alta eficiencia (HPLC).....	32
3.1.2.4 Estudios Moleculares.....	33
3.1.3 Clasificaciones de las anemias .....	35
3.1.3.1 Anemias adquiridas .....	36
3.1.3.2 Anemias congénitas.....	41
3.2 Introducción a la Inteligencia Artificial.....	47
3.2.1 Concepto e historia.....	47
3.2.2 Clasificación y breve introducción a los métodos de IA .....	51
3.2.2.1 Regresión lineal .....	53
3.2.2.2 Regresión Logística .....	54
3.2.2.3 Máquinas de Soporte Vectorial .....	55
3.2.2.4 Árboles de decisión.....	57
3.2.2.5 Redes neuronales artificiales .....	59
3.2.3 Aplicaciones de la Inteligencia Artificial en el mundo real .....	61

3.2.4 Aplicaciones de la Inteligencia Artificial en salud .....	63
3.3 Aplicaciones de Inteligencia Artificial en hematología .....	<b>68</b>
3.3.1 Inteligencia Artificial en hematología mediante el análisis de datos analíticos	71
3.3.2 Inteligencia Artificial en la clasificación de imágenes celulares.....	72
3.3.3 Inteligencia Artificial en la citometría de flujo.....	74
3.3.4 Inteligencia Artificial en aplicaciones clínicas de hematología .....	76
3.4 Inteligencia Artificial en la clasificación de anemias.....	<b>79</b>
3.4.1. Diagnóstico de anemias mediante ML tradicional en el análisis de datos tabulados.....	80
3.4.2. Diagnóstico de anemias mediante Redes Neuronales en el análisis de datos tabulados.....	81
3.4.3. Diagnóstico de anemia mediante el análisis de imágenes .....	83
3.4.4. Otras formas de aplicación de IA en el estudio de anemias.....	84
3.5 Limitaciones de la Inteligencia Artificial en salud.....	<b>86</b>
<b>4. Hipótesis y objetivos.....</b>	<b>89</b>
<b>5. Materiales y Métodos .....</b>	<b>92</b>
5.1 Búsqueda bibliográfica.....	<b>92</b>
5.1.1 Estrategia de búsqueda .....	92
5.1.2 Base de datos PUBMED (National Librery of Medicine).....	94
5.1.3 Base de datos EMBASE .....	95

5.1.4 Base de datos WOS (Web Of Sciences).....	96
5.1.5 Base de datos Dialnet .....	96
5.1.6 Portal Regional de la BVS.....	96
5.1.7 Tesis Doctorales.....	96
5.2 Tipo de estudio y fuente de datos .....	<b>97</b>
5.3 Variables incluidas en el entrenamiento del algoritmo .....	<b>99</b>
5.4 Criterios de inclusión.....	<b>100</b>
5.5 Sistemas de diagnóstico hematológico .....	<b>101</b>
5.6 Criterios de exclusión: limpieza de datos. ....	<b>103</b>
5.7. Digitalización de la muestra.....	<b>105</b>
5.8. Procesamiento de datos: Enseñando a la máquina.....	<b>107</b>
5.9. Métodos de Inteligencia Artificial aplicados .....	<b>108</b>
5.10. Experimentos realizados .....	<b>111</b>
5.11. Métodos de validación y explicabilidad del sistema de IA .....	<b>112</b>
5.12. Métricas utilizadas y conceptos asociados .....	<b>113</b>
5.13. Aspectos éticos .....	<b>115</b>
<b>6. Resultados.....</b>	<b>116</b>
6.1. Distribución de las muestras utilizadas para el entrenamiento y validación del sistema .....	117
6.2 Resultados finales.....	118

6.3 Resultados con datos vírgenes (sin limpieza de datos) .....	121
6.4 Curvas ROC y matrices de confusión relativas a los experimentos con BloodNet .	122
6.5 Resultados de explicabilidad utilizados sobre BloodNet.....	127
6.5.1 Gráficos de explicabilidad en el escenario 2: Pacientes sanos .....	128
6.5.2 Gráficos de explicabilidad en el escenario 2: Pacientes con anemias adquiridas .....	129
6.5.3 Gráficos de explicabilidad en el escenario 2: Pacientes con anemias congénitas (talasemias). .....	130
6.5.4 Gráficos de explicabilidad en el escenario 3: pacientes sanos .....	131
6.5.5 Gráficos de explicabilidad en el escenario 3: Anemias ferropénicas .....	132
6.5.6 Gráficos de explicabilidad en el escenario 3: Anemias por trastorno crónico.	133
6.5.7 Gráficos de explicabilidad en el escenario 3: Talasemias .....	134
6.5.8 Gráficos de explicabilidad en el escenario 3: Hemoglobinopatías estructurales .....	135
6.6 Resultados de la puesta a punto o validación externa de BloodNet.....	136
<b>7. Discusión .....</b>	<b>137</b>
7.1 El valor de la Inteligencia Artificial en Medicina. Evaluación de resultados.....	137
7.2 El valor de la calidad y cantidad de datos .....	142
7.3 El valor de aportar conocimiento específico a una Red Neuronal .....	146
7.4 La importancia de la explicabilidad en ciencias de la salud. La caja negra ( <i>The     Blackbox</i> ) .....	148

7.5 La importancia de la validación externa.....	150
7.6 La Importancia de las variables incluidas en el estudio .....	151
7.7 Comparando nuestros resultados con lo publicado previamente .....	153
7.8 Evaluación coste-efectiva.....	157
7.9 Limitaciones de esta tesis doctoral .....	158
7.10 Reflexión personal.....	160
<b>8. Conclusiones .....</b>	<b>162</b>
<b>9. Trabajos y ponencias en congresos que apoyan esta tesis doctoral.</b>	<b>165</b>
<b>10. Bibliografía .....</b>	<b>166</b>

## Índice de tablas

- Tabla 1. Resultados, clasificación y selección de la búsqueda bibliografía realizada.  
.....93
- Tabla 2. Población estudiada, dividida por centros de salud dependientes del Hospital Clínico San Carlos .....98
- Tabla 3. Muestras incluidas en el aprendizaje de BloodNet .....98
- Tabla 4. Variables y valores referencia utilizados por el analizador *Beckman Coulter-DxH*.....99
- Tabla 5. Diagnósticos divididos por subgrupos de las muestras de hemoglobinopatías incluidas en el entrenamiento del algoritmo inteligente .....100
- Tabla 6. Diagnósticos de hemoglobinopatías llevados a cabo en el Hospital Clínico San Carlos de Madrid en el año 2019.....102
- Tabla 7. Escenarios de cada uno de los entrenamientos realizados. ....111
- Tabla 8. Distribución de las muestras recogidas para el entrenamiento, validación y puesta a punto de la Red Neuronal BloodNet.....117
- Tabla 9. Resultados obtenidos en el escenario 1 .....118
- Tabla 10. Resultados obtenidos en el escenario 2 .....119
- Tabla 11. Resultados obtenidos en el escenario 3. ....120
- Tabla 12. Resultados obtenidos en el escenario 3.5 (datos vírgenes).....121
- Tabla 13. Resultados obtenidos de la puesta a punto o validación externa de BloodNet .....136
- Tabla 14. Resumen de la Exactitud/Accuracy de los experimentos realizados.....139

## Índice de Figuras

- Figura 1. Número de publicaciones científicas en PudMed entre 2002 y 2022 utilizando como búsqueda “(Artificial intelligence) AND medicine” .....24
- Figura 2. Resultados mostrados por GoogleTrends sobre “Inteligencia Artificial” ..25
- Figura 3. Semiología eritrocitaria. A: Drepanocitosis, B: Punteado basófilo, C)Dianocitos, D) Esferocitos e Hipocromía.....31
- Figura 4. Clasificación de las anemias por origen congénito y adquirido y por su mecanismo de acción hemolítico .....36
- Figura 5. Semiología eritrocitaria. Esquistocitos .....40
- Figura 6. Semiología eritroide. A: Eliptocitosis. B: Ovalocitosis. C: Esferocitosis .....42
- Figura 7. Clasificación general de la Inteligencia Artificial..... 51
- Figura 8. Clasificación de los diferentes modelos de Inteligencia Artificial que se tratan en esta tesis doctoral. ....52
- Figura 9. Representación de la regresión lineal y regresión logística.....54
- Figura 10. Ejemplo de hiperplano de 2 dimensiones .....55
- Figura 11. Representación del funcionamiento de la Máquina de Soporte Vectorial .....55
- Figura 12. Imagen representativa de la función de Kernel. ....56
- Figura 13. Ejemplo básico de Árbol de Decisión para el diagnóstico de anemias....57
- Figura 14. Representación esquemática de dos diferentes formas de crear un algoritmo basado en Arboles de Decisión .....58
- Figura 15. Ejemplo de Red Neuronal artificial, comparada con una neurona humana. ....59

- Figura 16. Flujo de trabajo entre el departamento de Hematología del HCSC y el departamento de arquitectura de la computación de la UMA .....105
- Figura 17. Interface del método de etiquetado durante esta tesis doctoral .....106
- Figura 18. Representación del proceso interactivo de datos en BloodNet .....109
- Figura 19. Matriz de confusión de Bloodnet en escenario 1 .....123
- Figura 20. Curvas ROC (izquierda) y PRC (derecha) de Bloodnet en escenario 1 ...123
- Figura 21. Matriz de confusión de Bloodnet en escenario 2 .....124
- Figura 22. Curvas ROC (izquierda) y PRC (derecha) de bloodnet en escenario 2 ...124
- Figura 23. Matriz de confusión de bloodnet en escenario 3 .....125
- Figura 24. Curvas ROC (izquierda) y PRC (derecha) de bloodnet en escenario 3 ...126
- Figura 25. Gráficos de explicabilidad para pacientes sanos (escenario 2) mediante Bloodnet. ....128
- Figura 26. Gráficos de explicabilidad para pacientes con anemias adquiridas (escenario 2) mediante Bloodnet. ....129
- Figura 27. Gráficos de explicabilidad para pacientes con Talasemias (escenario 2) mediante BloodNet. ....130
- Figura 28. Gráficos de explicabilidad para pacientes sanos (escenario 3) mediante BloodNet. ....131
- Figura 29. Gráficos de explicabilidad para pacientes con anemia ferropénica (escenario 3) mediante BloodNet. ....132
- Figura 30. Gráficos de explicabilidad para pacientes con anemia por trastorno crónico (escenario 3) mediante BloodNet.....133
- Figura 31. Gráficos de explicabilidad para pacientes con Talasemias (escenario 3) mediante BloodNet. ....134

- Figura 32. Gráficos de explicabilidad para pacientes con Hemoglobinopatías Estructurales (escenario 3) mediante BloodNet. .... 135
- Figura 33. Matriz de confusión de la validación externa o puesta a punto de BloodNet ..... 136
- Figura 34. Representación de la multidisciplinariedad en la ciencia del siglo XXI.. 138
- Figura 35. Matriz de confusión de la Red Neuronal (escenario 3) ..... 144
- Figura 36. Representación ilustrativa del procesamiento de datos por Bloodnet. 146
- Figura 37. Representación de uso real de BloodNet ..... 149
- Figura 38. Matriz de correlación de las variables incluidas en el estudio de esta tesis doctoral ..... 151

# 1. Resumen

**Título:** Medicina de precisión en hematología: Inteligencia Artificial para el diagnóstico diferencial de anemias.

## **Introducción**

La anemia es una de las afecciones más frecuentes a nivel mundial y según los últimos datos publicados por la Organización Mundial de la Salud (OMS), afecta al 24,8 % de la población mundial (1,62 billones de personas). Por otro lado, la Inteligencia Artificial (IA), una disciplina perteneciente al campo de las ciencias de la computación, ha llegado a nuestras vidas para formar parte de ellas en forma de aplicaciones de nuestro teléfono inteligente, reconocimiento facial, traducciones simultáneas, recomendaciones internautas e incluso en forma de prototipos de coches autónomos. Durante los últimos 10 años la aparición de la IA en medicina y hematología es evidente y de ella se esperan mejoras para el paciente, el médico y el sistema sanitario. Entre sus objetivos encontramos la reducción en los tiempos de espera, la minimización de los errores humanos y mejorar la eficiencia en la gestión de recursos. Sin embargo, el desconocimiento en cuanto a la explicabilidad de estos sistemas inteligentes sigue siendo una caja negra difícil de decodificar y mantiene escéptico a una parte del mundo sanitario.

## **Objetivos**

El principal objetivo de esta tesis doctoral es abordar la posibilidad real del uso de sistemas de Inteligencia Artificial para apoyar el diagnóstico diferencial de anemias. Para

ello desarrollaremos una Red Neuronal que será entrenada a partir de hemogramas, tratando de disminuir el tiempo de diagnóstico y los recursos utilizados sin perder precisión en el diagnóstico. Entre los objetivos secundarios encontramos: ofrecer una visión holística de la aplicación de IA en hematología, demostrar la importancia que tiene la calidad de los datos en el entrenamiento del sistema, demostrar mediante sistemas de explicabilidad el razonamiento del diagnóstico llevado a cabo por nuestra Red Neuronal, validar nuestros resultados en una segunda muestra independiente y comparar nuestra Red Neuronal con los sistemas inteligentes publicados en la bibliografía.

## **Materiales y métodos**

Para llevar a cabo esta tesis doctoral hemos creado un grupo de trabajo entre hematólogos del Hospital Universitario Clínico San Carlos de Madrid (HCSC) e informáticos expertos en IA de la Facultad de Ingeniería de la Computación de la Universidad de Málaga (UMA).

Hemos recogido más de 100.000 hemogramas procesados en el HCSC entre los años 2011 y 2018 y, a partir de ellos, hemos creado un conjunto de datos etiquetando 4.060 muestras con 4 diagnósticos diferentes: anemia ferropénica, anemia por trastorno crónico, talasemia y hemoglobinopatía estructural, además de añadir hemogramas de pacientes sanos. Las etiquetas han sido asignadas por un hematólogo mediante la utilización de una interface web diseñada para esta tesis doctoral que ha sido denominada *Labeller*. Estos datos etiquetados han servido como base para entrenar distintos modelos inteligentes, entre los que se encuentra nuestro diseño de Red

Neuronal (BloodNet) a la que hemos aportado conocimiento específico. Se ha realizado el entrenamiento sobre varios escenarios diferentes, desde el más sencillo (escenario 1) en el que se propone una clasificación binaria entre sanos y enfermos, hasta el experimento más complejo (escenario 3) que separa todos los tipos de anemias específicamente. Además, hemos realizado un análisis extra al ejercicio más complejo sin llevar a cabo ninguna limpieza de datos (escenario 3.5) con lo que pretendemos demostrar la importancia en la calidad de datos de la muestra.

Hemos aplicado modelos de explicabilidad sobre los resultados obtenidos mediante el uso de valores de SHAP y gráficos de Beeswarm. Además, hemos llevado a cabo una puesta a punto o validación externa con una segunda muestra de 250 hemogramas.

## **Resultados**

Los resultados obtenidos muestran de manera global un *Accuracy* o Exactitud del 88,6% a la hora de clasificar correctamente 4 de las 5 entidades propuestas y su área bajo la curva ROC ha alcanzado valores entre el 0,97 y 0,98 dependiendo de la categoría diagnosticada. Las características intrínsecas de las hemoglobinopatías estructurales, así como la necesidad de una mayor muestra han hecho que no se alcancen resultados óptimos para el diagnóstico de esta entidad.

En cuanto a los resultados de los escenarios de menor complejidad (sanos vs enfermos) la exactitud de nuestra Red Neuronal alcanza el 96,4%. Por otro lado, el experimento realizado sobre una muestra sin filtros de limpieza de datos (escenario 3.5) ha mostrado una exactitud del 56%.

Los sistemas de explicabilidad han determinado que las variables Hemoglobina Corpuscular Media, Hemoglobina, Hematocrito, Número de hematíes y Volumen Corpuscular Medio son las de mayor peso a la hora de clasificar las anemias; siendo los Leucocitos totales la variable de menor peso.

Los resultados obtenidos se han ratificado en una muestra segunda independiente de 250 hemogramas alcanzándose un Accuracy del 90% en el escenario más complejo.

Una vez que el sistema inteligente está diseñado, el diagnóstico se obtiene de manera instantánea a partir del hemograma, siendo éste su único coste.

## **Conclusiones**

Nuestra Red Neuronal (BloodNet) ha conseguido una exactitud, precisión, sensibilidad y área bajo la curva ROC similar o mayor a los publicados en el uso de la Inteligencia Artificial para el diagnóstico de anemias, pese a la mayor complejidad del problema planteado en esta tesis doctoral. BloodNet ha mostrado resultados superiores al del resto de sistemas de Machine Learning comparados (incluida una Red Neuronal de uso público denominada TabNet).

Las hemoglobinopatías estructurales han supuesto el mayor problema diagnóstico, dada la variabilidad entre sus subtipos y su baja expresión en el hemograma. Obtener datos de alta calidad debe ser un objetivo principal donde centrar el inicio de la investigación en proyectos de Inteligencia Artificial y medicina, ya que afecta claramente a la exactitud del algoritmo inteligente.

Los datos de explicabilidad muestran que el razonamiento de nuestra maquina inteligente es similar al de la práctica clínica. Además, hemos confirmado la validez de nuestra Red Neuronal en una segunda muestra independiente a la utilizada para el entrenamiento del sistema. Todo ello aporta confianza en el sistema y aumenta su valor en la práctica habitual.

Hemos reducido el tiempo y los costes sin disminuir la precisión del diagnóstico. Además, BloodNet facilita la accesibilidad al diagnóstico a la mayoría de centros de atención médica, al utilizar hemogramas únicamente como fuente de diagnóstico. Todo ello podría posicionar a BloodNet como una mejora en la gestión de recursos para el diagnóstico diferencial de anemias.

## 2. Abstract

**Title:** Precision medicine in hematology: artificial intelligence for the differential diagnosis of anemias.

### **Introduction**

Anemia is one of the most common conditions worldwide and according to the most recent data published by the World Health Organization (WHO), it affects 24.8% of the world's population (1.62 billion people). On the other hand, artificial intelligence (AI), a discipline belonging to the field of computer science, has arrived in our lives to become part of them in the form of smartphone applications, facial recognition, simultaneous translations, movie and song recommendations and even like autonomous car prototypes. Over the last 10 years the emergence of AI in medicine is also evident and improvements for the patient, the physician and the healthcare system are expected. Its objectives include short lead times, minimizing human error and improve resource management. However, the explainability of these intelligent systems remains a black box that is difficult to decode and keeps part of the healthcare world skeptical.

### **Objectives**

The main objective of this doctoral thesis is to improve the differential diagnosis systems of anemias through the development of a neural network trained only with hemograms,

trying to decrease the diagnostic time and resources used without decreasing the diagnostic accuracy.

Secondary objectives include: to offer a holistic view of the application of AI in hematology, to demonstrate the importance of the quality of the data collected, to show the reasoning behind the learning of our neural network by means of explainability systems, and to compare our neural network with systems published in the literature.

## **Materials and methods**

To carry out this PhD thesis we have created a multidisciplinary working group between hematologists and computer scientist experts in AI from the Hospital Universitario Clínico San Carlos de Madrid (HCSC) and the Faculty of Computer Engineering of the University of Malaga (UMA) respectively.

We have collected more than 100,000 hemograms processed at HCSC between the years 2011 and 2018 and we have created a dataset labeling 4,060 samples with 4 different diagnoses: iron deficiency anemia, anemia due to chronic disorder, hemoglobinopathies and hemograms of healthy patients. The labels have been assigned to each of the hemograms by a hematologist using a web interface designed for this doctoral thesis, which has been named Labeller. These labeled data have been the basis for the training of different intelligent models, including our neural network design with specific knowledge (BloodNet). Training has been performed on several different stages, from the simplest (stage 1) in which a binary classification between healthy and sick is proposed, to the most complex experiment (stage 3) that separates all types of anemias specifically. In addition, we have performed an extra analysis to the most complex

exercise without performing any data cleaning (stage 3.5) in order to demonstrate the importance of the data quality of the sample.

We have applied explanatory models using SHAP values and Beeswarm plots. In addition, we have performed an external validation with a second sample of 250 blood counts.

## **Results**

The results obtained show an overall Accuracy of 88.6% in correctly classifying 4 of the 5 proposed diseases and their area under the ROC curve has reached values between 0.97 and 0.98 depending on the diagnosed category. Intelligent systems compared with our neural network have been shown to be inferior (including a publicly available neural network called TabNet).

When the results of scenarios with low complexity (healthy vs sick) are evaluated, the accuracy of our neural network increases to 96.4%. The experiment performed on a sample without data cleaning filters has shown an accuracy of 56%.

The explainability systems have determined that the variables Mean Corpuscular Hemoglobin, Hemoglobin, Hematocrit, Number of red blood cells and Mean Corpuscular Volume are the most important variables when classifying anemias.

The results obtained have been ratified in a second independent sample of 250 hemograms, reaching an Accuracy of 90% in the most complex scenario.

Once the intelligent system is designed, the diagnosis is obtained instantaneously from the hemogram, this being its only cost.

## **Conclusions**

Our neural network (BloodNet) has achieved accuracy, precision, sensitivity and an area under the ROC curve similar to or greater than the results published in the scientific literature, despite the greater complexity proposed in this doctoral thesis, as it includes four diagnoses of anemias. Structural hemoglobinopathies have been the most difficult diseases to classify, given the variability between their subtypes and their low expression in the hemogram. We have reinforced the idea that the quality of the data obtained should be one of the main objectives on which to focus the beginning of the research, since it directly affects the accuracy of the intelligent system. The explainability data show that the reasoning of our intelligent machine is similar to hospital practice, which gives confidence in the system and increases its value. Our neural network has shown superiority over other neural networks and traditional Machine Learning algorithms compared. We have validated our Neural Network with a second sample independent of the one used for training (external validation). The time from blood sample collection to diagnosis has decreased dramatically from weeks to seconds. Performing the differential diagnosis based only on a hemogram facilitates its accessibility to most health care centers. This, together with its low cost, makes this system an improvement in the management of resources for the differential diagnosis of anemias.

### 3. Introducción

La anemia es una de las afecciones más frecuentes a nivel mundial. Según los últimos datos publicados por la Organización Mundial de la Salud (OMS), la prevalencia global es de 24,8 %, afectando a 1,62 billones de personas en el mundo (1), aunque existen publicaciones más recientes que aumentan este porcentaje hasta el 32,9 % (2). Además, esta prevalencia puede ascender dramáticamente en poblaciones concretas, siendo de hasta el 47,4 % en niños preescolares y al 41,8 % en mujeres embarazadas(3). Por tanto, su correcto y precoz diagnóstico debe estar presente como objetivo primordial para la OMS y los países que la integran, entre los que se encuentra España.

El gasto económico en recursos materiales y humanos necesarios para el correcto diagnóstico hasta alcanzar la verdadera etiología de la anemia comprende un abanico muy heterogéneo de posibilidades. Tanto es así que los recursos utilizados pueden ir desde una simple analítica con hemograma y bioquímica hasta la necesidad de realización de estudios citogenéticos y de biología molecular (frecuentemente utilizados para el diagnóstico de hemoglobinopatías y otras anemias de origen congénito). En este contexto, es importante recordar que las hemoglobinopatías son la alteración monogénica más frecuente del mundo, conociéndose que más de 5 % de la población mundial porta alguna alteración genética asociada (4). De hecho, en 2008 la OMS catalogó a las hemoglobinopatías como un problema de salud pública (5).

Por otro lado, la Inteligencia Artificial (IA), una disciplina perteneciente al campo de las ciencias de la computación que comprende desde algoritmos sencillos hasta el uso de complejos sistemas basados en los sistemas neuronales humanos, ha llegado a nuestras

vidas para formar parte de ella. Durante nuestro día a día somos consumidores activos de ella con el uso de nuestros *Smartphones*, siguiendo las sugerencias de búsquedas internautas, las recomendaciones para disfrutar una película, a la hora de escuchar un disco propuesto por la IA que se adapta a nuestros gustos o participando en la selección de la próxima publicación que nuestra red social va a mostrarnos. Aunque estas sean las formas de IA con las que más nos relacionamos, el progresivo crecimiento de esta tecnología abarca todos los campos que podamos imaginar, entre los cuales encontramos los subtítulos autogenerados, las traducciones simultáneas, el transporte y gestión de mercancías, los prototipos de coches autónomos o incluso proyectos tan ambiciosos como las ciudades inteligentes(6,7).

En ciencias de la salud, la aplicación de diferentes formas de inteligencia es también una realidad desde hace más de 20 años, encontrándose publicaciones a este respecto de hace hasta 40 años donde ya se habla del binomio anemia e Inteligencia Artificial (8). No obstante, debemos tener en cuenta que el concepto de IA es muy amplio y ha ido evolucionando, llegando a ser utilizado como sinónimo de sistemas computacionales en

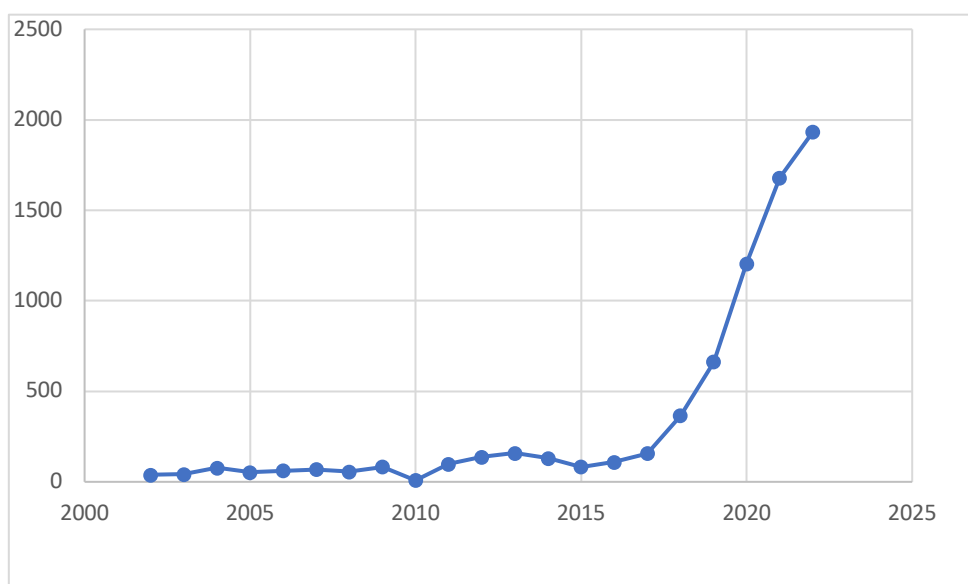


FIGURA 1. NÚMERO DE PUBLICACIONES CIENTÍFICAS EN PUDMED ENTRE 2002 Y 2022 UTILIZANDO COMO BÚSQUEDA “(ARTIFICIAL INTELLIGENCE) AND MEDICINE”

sus momentos más iniciales. Pese a estas publicaciones tan visionarias de antaño, no es hasta la última década (incluso podríamos decir que hasta el último lustro), cuando se produce una verdadera explosión de la IA en el campo de la medicina, como vemos reflejado en la **Figura 1**. Con la aparición de la misma se esperan mejoras tanto para el médico, como para el sistema sanitario y por supuesto para el paciente. Entre sus objetivos marcados encontramos la mejoras en los tiempos de espera, la disminución de los errores médicos y la mejora en la gestión de los recursos disponibles (9–12), siendo el objetivo principal el de ofrecer una medicina muy personalizada y de alta precisión (13).

El actual crecimiento del binomio IA-Medicina ha sido posible gracias a las grandes cantidades de datos generadas a partir de la informatización de los sistemas sanitarios y a los avances en cuanto a su almacenamiento y manejo (Big Data y minería de datos), que hasta entonces habían sido un freno del proceso (14).

Para situarnos en el carácter actual de este proyecto podemos ver como la reciente publicación de La Fundación del Español Urgente (FundéuRAE), promovida por la Agencia EFE y la Real Academia Española, ha determinado que “*Inteligencia Artificial*” sea la expresión-palabra del año 2022, superando a *criptomoneda*, *ucraniano* o *inflación*, entre otras (15). Así mismo, vemos que el interés mundial generado en IA durante los últimos 5 años está en sus máximos, como se muestra en la **Figura 2**.



FIGURA 2. RESULTADOS MOSTRADOS POR GOOGLE®TRENDS SOBRE “INTELIGENCIA ARTIFICIAL” A 27 DE ABRIL DEL 2023.

Este proyecto de tesis doctoral aborda la posibilidad real del uso de sistemas de Inteligencia Artificial para apoyar el diagnóstico diferencial de anemias. En concreto, diseñaremos una Red Neuronal propia y capaz de conseguir un diagnóstico de precisión entre las anemias más frecuentes, a partir de un único hemograma. Compararemos nuestros resultados con otras técnicas de *Machine Learning (ML)* tradicional, tales como los Árboles de Decisión (DT, del inglés *Decision Trees*) o las Máquinas de Soporte Vectorial (*SVM, del inglés Support Vector Machine*), además de compararlos con los de otras Redes Neuronales pertenecientes al estado del arte. Todo ello con la única tarea de ayudar (y no sustituir) al sanitario en el diagnóstico de las anemias adquiridas y congénitas más frecuentes.

Además, trataremos de mejorar los resultados publicados por otros grupos mediante la inclusión de conocimiento clínico específico del caso de uso en la propia Red Neuronal. Nuestro objetivo es entrenar un algoritmo inteligente usando las diferentes variables del hemograma como único dato de entrada, además de la edad y el sexo, lo que facilitará su accesibilidad a todos los centros de atención sanitaria, ya que no todos los centros disponen de estudios moleculares para el diagnóstico de hemoglobinopatías, por ejemplo.

Para ello hemos creado un grupo de trabajo multidisciplinar entre médicos e informáticos expertos en IA de diferentes hospitales y universidades (Madrid y Málaga respectivamente). Tras la recolección de más de 100.000 hemogramas entre el año 2011 y 2018 del Hospital Clínico San Carlos de Madrid, hemos creado un conjunto de datos etiquetados con diferentes diagnósticos de anemias de 4.060 muestras que han servido como base para entrenar distintos modelos inteligentes, entre los que se encuentra

nuestro diseño de Red Neuronal con conocimiento específico que hemos denominado BloodNet.

Asimismo, hemos centrado gran parte de nuestros esfuerzos en entender el funcionamiento de nuestra Red Neuronal mediante sistemas de explicabilidad y de darle validez externa con el fin de que sea acogida por el gremio sanitario como una aplicación fiable para el apoyo en el diagnóstico diferencial de anemias.

## 3.1 Introducción a las anemias

### 3.1.1 Concepto, importancia y prevalencia

La Hemoglobina (Hb) es una proteína globular formada por 4 cadenas de globinas, que forma parte indispensable en el hematíe y el hierro que transporta en su interior le aporta al hematíe y a la sangre su color rojo característico. Su función transportadora de oxígeno (O<sub>2</sub>) desde los pulmones hacia los tejidos y a su vez del dióxido de carbono (CO<sub>2</sub>) desde los tejidos a los pulmones, es vital para la supervivencia celular, siendo así uno de los eslabones básicos de la pirámide donde se sustenta nuestra supervivencia. La disminución de la Hb en dos desviaciones estándar con respecto a la media de los valores poblacionales normales (ajustado a edad y sexo) es lo que denominamos anemia (4,16,17).

La anemia no se considera una enfermedad *per se*, sino una forma de presentación de una patología en la que, como consecuencia de la misma: a) el paciente no es capaz de producir hematíes correctamente; b) los hematíes del paciente se destruyen (hemólisis); c) los hematíes del paciente se pierden en mayor cantidad a la habitual (sangrados). Las diferentes causas de anemia y su clasificación serán objeto de un apartado de esta tesis.

Además de su altísima presentación a nivel global, en torno al 30 %, el interés de las anemias radica en que puede ser la única forma de presentación de muchas enfermedades (2,18,19). La causa más frecuente de anemia es el déficit de hierro o ferropenia, pero existen también anemias de origen genético cuya prevalencia es elevada en nuestro medio nacional (España), como las hemoglobinopatías.

Las hemoglobinopatías a nivel mundial se encuentra presentes en el 5% de la población, siendo algo mayor en nuestro medio (5). Según las últimas publicaciones, la prevalencia

en España se sitúa en un 6-8% de la población, tratándose la mayoría de ellas de talasemias (4,89- 6,89%), las cuales cumplen una distribución muy heterogénea (20). Cuando nos referimos a hemoglobinopatías estructurales, su tasa de presentación aún se sitúa en el 1-2% pero, debido al creciente flujo migratorio, la incidencia se encuentra en progresivo crecimiento, como se demuestra en diversos trabajos realizados en nuestro país (20–22), como en otros puntos de Europa (23).

La elevada prevalencia, tanto de anemias congénitas como adquiridas, nos obliga a considerar la anemia como un problema de salud pública nacional e internacional, algo que ya anunció la OMS en 2008 (5).

En los países desarrollados se dispone desde hace años de opciones terapéuticas suficientes para evitar que la supervivencia y calidad de vida de estos pacientes esté afectada por la anemia, pero existen dificultades para poder afirmar lo mismo en países en vías de desarrollo o subdesarrollados, llegando las hemoglobinopatías a considerarse causa de un 3,4% de las muertes anuales en menores de 5 años(5,24).

### 3.1.2 Bases del diagnóstico actual de las anemias

El diagnóstico de las anemias, así como sus posibles orientaciones etiológicas, está basado en el resultado inicial del hemograma, realizado a partir de una muestra sanguínea. Posteriormente tendremos que realizar una batería de pruebas sucesivas cuya realización se irá basando en el resultado de la anterior, hasta conocer la causa de la anemia (25).

#### 3.1.2.1 El hemograma

El hemograma es una prueba rutinaria básica y económica que está presente en la mayoría de valoraciones médicas, tanto ambulatorias como hospitalarias. El hemograma nos muestra datos cuantitativos y cualitativos de los elementos formes del torrente sanguíneo (leucocitos, plaquetas y eritrocitos). Por tanto, además de mostrarnos la anemia (desviación a la baja de la cifra de Hemoglobina 2 desviación con respecto a la media poblacional), el hemograma nos dará información acerca del tamaño del hematíe, la concentración de Hemoglobina en cada uno de ellos, la diferencia de tamaño entre los hematíes de una misma muestra, etc. Además, la realización del hemograma se suele acompañar de un estudio bioquímico de la misma muestra, con la cual podremos orientar mejor la causa de la anemia y, si lo precisa, dirigir el estudio complementario y el posible tratamiento (26).

#### 3.1.2.2. El frotis de sangre periférica

De la misma manera, la descripción morfológica de los glóbulos rojos mediante la visualización óptica de la extensión de sangre periférica (o frotis de sangre periférica)

nos ayudará a orientar la etiología de la anemia, existiendo morfologías eritroides características de ciertas enfermedades (incluso patognomónicas en algunos casos). Por

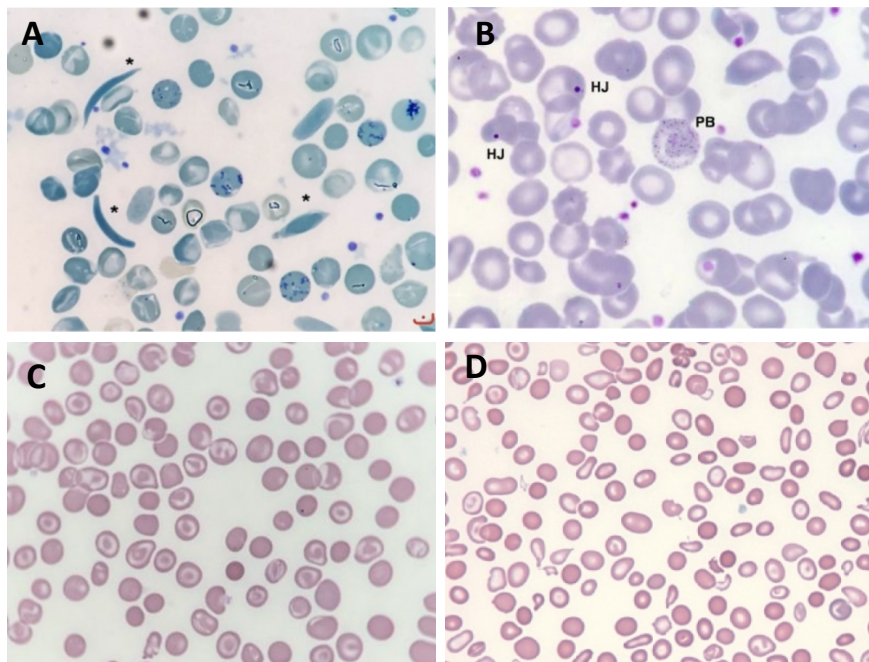


FIGURA 3. SEMIOLOGÍA ERITROCITARIA. A: DREPANOCITOSIS. B: PUNTEADO BASÓFILO (PB). C)DIANOCITOSIS. D) ESFEROCITOS E HIPOCROMÍA. IMÁGENES TOMADA DEL ATLAS DEL GRUPO ESPAÑOL DE CITOLOGÍA HEMATOLÓGICA.

ejemplo, los drepanocitos o células falciformes son patognomónicos de la anemia drepanocítica o Hemoglobinopatía S homocigota. De la misma manera, los esferocitos son característicos de las esferocitosis hereditarias y anemias hemolíticas autoinmunes, el punteado basófilo de las Beta-talasemias y la microcitosis y la hipocromía la podemos ver tanto en talasemias como en anemias ferropénicas, como vemos en la **Figura 3** (27). Aunque normalmente las talasemias tienen mayor componente de microcitosis y la ferropenia de hipocromía, esta característica compartida puede llevar a confundir la orientación diagnóstica entre ambas, algo que comentaremos repetidamente a lo largo de esta tesis, por su importancia. Por esto, la extensión de sangre periférica es considerada un componente imprescindible en la batería de pruebas a realizar en un estudio de anemia(28).

Cuando tras la realización de las pruebas iniciales (hemograma, bioquímica y frotis de sangre periférica) hemos descartado razonablemente las anemias más frecuentes (adquiridas) y pensamos estar frente a una anemia de origen congénito, necesitamos ampliar una serie de pruebas, entre las que se encuentran la electroforesis capilar y la cromatografía líquida de alta eficacia (HPLC de sus siglas en inglés *high performance liquid chromatography*). Finalmente, dependiendo de los resultados obtenidos decidiremos qué estudio molecular (genético) necesitamos realizar: PCR (reacción en cadena de la polimerasa), MLPA (Amplificación de Sonda dependiente de ligandos) o secuenciación directa del gen en cuestión. Estos estudios son considerados el *Gold standard* en el diagnóstico de estas entidades. Estas técnicas genéticas tienen costes elevados y debemos elegir cuidadosamente cual escogemos. Para elegir bien debemos tener presente el volumen de muestras diarias que se van a procesar, el tipo de muestras, la disponibilidad de las técnicas en cada laboratorio, los costes y la experiencia del personal (26) .

Dado que en esta tesis se incluyen únicamente hemoglobinopatías y no otro tipo de anemias congénitas, describiremos únicamente cómo es el flujo de trabajo en un laboratorio con experiencia en diagnóstico de hemoglobinopatías (talasemias y hemoglobinopatías estructurales).

### 3.1.2.3 Electroforesis capilar y cromatografía líquida de alta eficiencia

(HPLC)

La electroforesis capilar (EC) y la cromatografía capilar de alta eficiencia (HPLC, de sus siglas en inglés: *High-Performance Liquid Chromatography*) son las técnicas más

sensibles y precisas para el diagnóstico de hemoglobinopatías estructurales y talasemias. Además son técnicas automatizadas, frente a la electroforesis tradicional, lo que contribuye de manera importante a la disminución de errores de realización e interpretación (29,30).

De manera general, ambas son técnicas que tienen el objetivo de separar moléculas con características diferentes a las que se observan en el paciente sano. La EC utiliza la diferencia de cargas para separar las diferentes moléculas (en este caso Hemoglobinas) presentes en una disolución. La separación se lleva a cabo en un tubo hueco de diámetro muy pequeño que denominamos capilar. Según las cargas de cada Hemoglobina patológica, estas se desplazarán hacia el ánodo o cátodo, mostrándose frente a la Hemoglobina sana HbA1.

Por otro lado, el HPLC es una técnica que utiliza las características químicas de cada Hemoglobina para ser capaz de separarlas dentro de la columna cromatográfica. Estos componentes se separan a medida que van pasando a través de esta columna, dependiendo de las diferentes interacciones químicas entre las sustancias a analizar y las sustancias de la columna. La movilización de la sustancia a estudio (moléculas de Hb) a través de la columna se produce por la presión que se transmite a través de la columna.

#### 3.1.2.4 Estudios Moleculares

El estudio molecular es indispensable para confirmar el diagnóstico de las hemoglobinopatías estructurales y talasemias. Además, es necesario para identificar pacientes en fases iniciales o presintomáticas de la enfermedad (despistaje neonatal) en los que tenemos alta sospecha de enfermedad grave. El estudio molecular se puede

llevar a cabo realizando un estudio de mutaciones puntuales o mediante el estudio de deleciones.

- Mutaciones puntuales: La mayoría de esos estudios se realizan mediante PCR (reacción en cadena de la polimerasa), es decir, mediante la utilización de *Primers* mutados y no mutados para la alteración genética que se está buscando. Otra manera de estudiar mutaciones puntuales es mediante la secuencia directa del gen, con el fin de estudiar todas las mutaciones puntuales posibles, aunque solo se suele utilizar cuando la sospecha de Hemoglobinopatía es alta pero no hemos encontrado las mutaciones puntuales más frecuentemente descritas en la literatura.
- Deleciones: Aquí también disponemos de diferentes arsenales, como Southern Blot, PCR gap o el MLPA. En esta última se utilizan múltiples sondas que se puedan unir a zonas diferentes a lo largo de todo el clúster estudiado y solo se amplifican cuando se detecta la mutación(31).

El coste de todos estos estudios puede ser muy diferente, pero en todos los casos elevado, tanto si hablamos de coste monetario de los materiales utilizados como si nos referimos a los recursos humanos.

Como hemos visto, llegar a conocer la causa de una anemia puede ir desde un hemograma y una bioquímica cuyo coste no supera los 50€ hasta los extensos estudios genéticos, que pueden llegar a tener un coste de hasta 2.000€ por paciente (32,33).

### 3.1.3 Clasificaciones de las anemias

Las anemias pueden ser consecuencia de muy diversas causas, incluso pueden ser de causa multifactorial y no existe un consenso internacional único para su clasificación. Es por ello que, dependiendo del libro que estudiemos, la escuela en la que nos hayamos formado o el artículo científico que nos leamos, podemos encontrar cualquiera de las siguientes clasificaciones(4):

#### A. Clasificación por el tamaño del eritrocito:

- *Macrocíticas*, donde el tamaño medio del hematíe está aumentado.
- *Normocíticas*, donde el tamaño medio del hematíe es normal.
- *Microcíticas*, donde el tamaño medio del hematíe está disminuido.

#### B. Clasificación por su comportamiento:

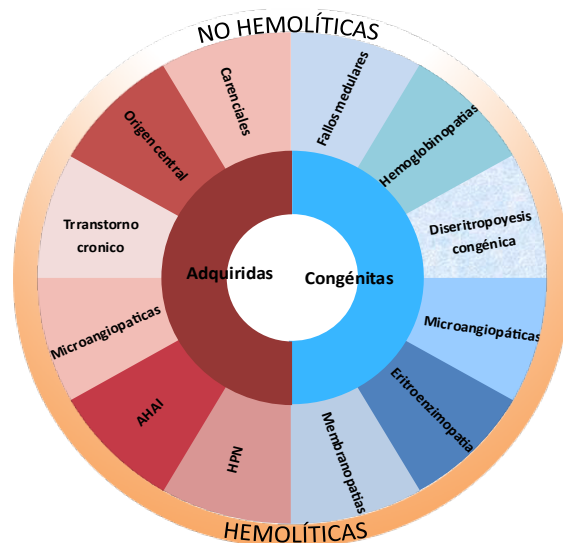
- *Hemolíticas*, en las que se produce la destrucción temprana del hematíe mediante diferentes mecanismos, disminuyendo su supervivencia y produciendo anemia en aquellos pacientes cuya producción de hematíes sea menor a la destrucción de los mismos (compensación insuficiente).
- No hemolíticas.

#### C. Clasificación por su origen:

- *Central*, donde existe una disminución en la producción en médula ósea de los progenitores eritroides.
- *Periférica*, donde existen pérdidas (sobre todo representadas en forma de sangrado) o una destrucción de los hematíes (hemólisis).

#### D. Clasificación por su forma de adquisición:

- *Congénitas*, donde el paciente es portador de una alteración genética desde su nacimiento y que acompañará durante toda su vida. Esta alteración suele asociar un cambio estructural o funcional del hematíe que acaba desembocando en mayor o menor medida en anemia (por dificultad en la producción medular o por hemólisis periférica).
- *Adquiridas*. Todas aquellas anemias adquiridas a lo largo de la vida.



**FIGURA 4.** CLASIFICACIÓN DE LAS ANEMIAS POR ORIGEN CONGÉNITO Y ADQUIRIDO Y POR SU MECANISMO DE ACCIÓN HEMOLÍTICO. IMAGEN PROPIA,

Esta última clasificación representada en la **Figura 4** es la que tiene mayor uso práctico y es la que vamos a desarrollar en las próximas páginas para contextualizar nuestro trabajo (4).

### 3.1.3.1 Anemias adquiridas

Las anemias adquiridas son con diferencia las más frecuentes y corresponden a más del 80% de todas las anemias a nivel mundial y a más del 90% de las anemias que valoramos en la práctica clínica de nuestro medio nacional (2). Como su nombre indica, se

adquieren a lo largo de la vida y son de muy diferente índole, siendo su mayor representante es la anemia de origen ferropénico.

#### 3.1.3.1.1 Anemias Carenciales: Anemia ferropénica

Como su nombre indica estas anemias son secundarias a la carencia o déficit de alguno de los substratos necesarios para la formación del hematíe. Entre ellas se encuentra la anemia por déficit de ácido fólico o vitamina B12, pero la más frecuente con diferencia es la anemia ferropénica (por déficit de hierro). Aunque no se puede determinar qué porcentaje del total de las anemias se deben únicamente al déficit de hierro, se han publicado estimaciones en las que ronda el 50-60%, dependiendo de la población estudiada(2,19). La anemia ferropénica puede estar causada por:

- Pérdidas de sangre: siendo los sangrados digestivos las causas más frecuentes en el varón y los ginecológicos en la mujer.
- Déficit nutritivo: en su mayoría representadas en países con recursos limitados.
- Hemólisis y el intento de reposición eritroide por parte de la médula, lo que llevará al consumo de las reservas de hierro.
- Malabsorción intestinal.

Las dos primeras son con diferencia las más frecuentes a nivel global, aunque dependiendo de si nos encontramos en un punto u otro del globo terráqueo sufriremos más una causa o la otra.

### 3.1.3.1.2 Anemias por trastorno crónico

Las anemias por trastorno crónico pueden desarrollarse en el contexto de un amplio abanico de enfermedades con un mayor o menor componente inflamatorio asociado, ya sea agudo o crónico y es la segunda causa más frecuente de anemia, siendo su prevalencia estimada de hasta un 40% del total de las anemias (34). Este proceso inflamatorio va a activar el sistema mononuclear-fagocítico y con ello, aumentarán los niveles de hepcidina. Mediante la regulación de la ferroportina, la elevación de la hepcidina desembocará en menor disponibilidad de hierro, apareciendo la anemia. Es decir, existen depósitos de hierro adecuados, pero el organismo no es capaz de realizar una buena gestión del mismo.

Además, la respuesta de la eritropoyetina no va a ser la adecuada ante la presentación de la anemia. Podemos decir que la eritropoyesis sufre un bloqueo secundario al proceso inflamatorio del paciente, que será mayor o menor dependiendo del contexto inflamatorio del paciente.

Entre las causas de anemia por trastorno crónico encontramos enfermedades infecciosas (sepsis, sarcoidosis, tuberculosis, etc.), enfermedades autoinmunes (lupus eritematosos sistémico, artritis reumatoide, etc.), así como enfermedades renales (mayor órgano productor de eritropoyetina a nivel fisiológico) o hepáticas.

### 3.1.3.1.3 Anemias de origen central

En este gran grupo de anemias adquiridas hemos incluido todas aquellas anemias por incapacidad o insuficiencia en la producción de glóbulos rojos desde la médula ósea. Aquí se encuentran representadas las enfermedades hematológicas en las que se

adquieren alteraciones genéticas que afectan a la eritropoyesis (Síndromes mielodisplásicos y Leucemias agudas en su mayoría). También se encuentran incluidas todas aquellas enfermedades hematológicas que producen anemia de manera secundaria al desplazamiento de la eritropoyesis medular resultado de la ocupación masiva medular (Linfomas y Mielomas múltiples como mayores representantes). Asimismo, las neoplasias de órgano sólido también pueden producir invasión medular (carcinomatosis medular), que van a producir el mismo efecto.

Para finalizar este grupo tan heterogéneo de anemias, no podemos olvidarnos de las aplasias medulares y aplasias pura de serie roja están representadas en este grupo. Enfermedades que se desarrollan de manera secundaria a una respuesta inmune patológica del organismo frente a diferentes posibles procesos y que acaba desembocando en una destrucción errónea de los precursores eritroides mediada por linfocitos (4).

#### 3.1.3.1.4 Anemias Hemolíticas

Este grupo de anemias adquiridas está representado por las anemias hemolíticas autoinmunes (AHAI), las anemias hemolíticas mecánicas, las anemias microangiopáticas y la Hemoglobinopatía Paroxística Nocturna (HPN), cada una de ellas con una fisiopatología diferente, pero con un desenlace similar: la destrucción anómala y aumentada del hematíe.

A) Anemia hemolítica autoinmune (AHAI): En ella se produce una destrucción del hematíe mediada por la aparición de anticuerpos que tendrán afinidad por los antígenos de membrana de los hematíes. Este acople del anticuerpo en la membrana

del hematíe, aumentará su destrucción por el sistema fagocítico en el bazo, desembocando en la aparición de anemia (35).

- B) Anemia hemolítica mecánica: Estas anemias son poco frecuentes y en su mayoría son secundarias a la función deteriorada de las prótesis mecánicas cardíacas y que, al paso de hematíe por éstas, lo destruirán por cizallamiento, observándose esquistocitos (hematíes rotos) en el frotis de sangre periférica (**Figura 5**).

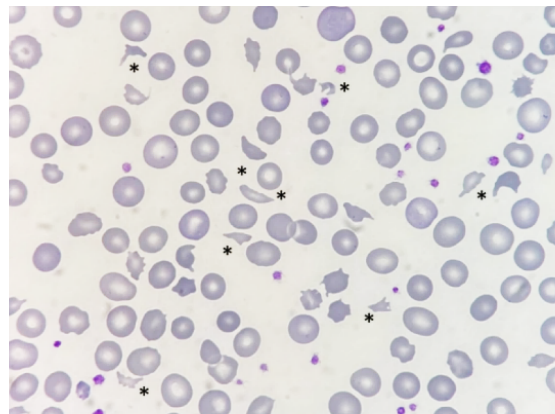


FIGURA 5. SEMIOLOGÍA ERITROCITARIA. \*ESQUISTOCITOS

- C) Anemia microangiopática trombótica (MAT): En este caso la destrucción del hematíe se producirá a nivel de la microcirculación capilar y puede estar mediada por múltiples causas como es el déficit de Adamst13 en la purpura trombocitopénica trombótica (PTT), los tumores sólidos en la coagulación intravascular diseminada (CID), por causas infecciosas como en el síndrome hemolítico urémico (SHU) e incluso por fármacos (36).
- D) Hemoglobinopatía paroxística nocturna (HPN): La fisiopatología de esta enfermedad está basa en la ausencia de producción de una proteína de anclaje en la membrana del hematíe llamada Glucosil Fosfadidil Inositol (GPI), secundaria a una alteración genética adquirida. La ausencia de GPI impide el anclaje de CD55 y CD59 a la membrana del hematíe, entre otras proteínas. La ausencia de expresión de CD59 y

CD55 impedirá la inhibición de la cascada del complemento, que marcará el hematíe como célula objetivo a destruir (37).

### 3.1.3.2 Anemias congénitas

Por definición este grupo de anemias acompañaran al paciente desde el nacimiento hasta su fallecimiento, exceptuando algunos pacientes que se someten a tratamientos curativos con indicaciones muy limitadas (como el trasplante medular o la terapia génica en Beta-talasemias). Por tanto, serán trasmisibles a descendencias, según el patrón de herencia correspondiente. La prevalencia mundial es muy heterogénea dependiendo de la zona geográfica a la que nos refiramos.

Las hemoglobinopatías son las mayores representantes de este grupo, y se estima que hasta el 20-25% de la población mundial es portador de alguna alteración genética asociada, llegando a alcanzar prevalencias de hasta el 45% en el sudeste asiático o hasta el 80% de algunas zonas muy endémicas (5). Pese a estos números tan llamativos, la mayoría de estas alteraciones son leves y asintomáticas. Por tanto, su diagnóstico debe centrarse en el consejo genético de cara a una futura descendencia.

#### 3.1.3.2.1 Anemias por alteración de la membrana del hematíe

Las alteraciones de la membrana de los hematíes o membranopatías son enfermedades genéticas hereditarias (en su grandísima mayoría) en las que, debido a una alteración en uno o varios genes que codifican para proteínas que conforman la membrana del hematíe, van a producir desestructuraciones o pérdidas de características de la membrana del hematíe. Estas propiedades perdidas van a impedir al hematíe que se

deforme correctamente para poder llegar a todos los capilares del sistema arteriovenoso.

Otras alteraciones genéticas que se pueden producir van a impedir la correcta permeabilidad de la membrana, lo que conlleva una menor supervivencia.

Todas estas patologías desembocan en el aumento de la destrucción del hematíe por hemólisis y, por tanto, en la aparición de anemia (en aquellos pacientes que no sean capaces de producir glóbulos rojos suficientes para compensar esta pérdida). Estas anemias van a adquirir el nombre de la alteración morfológica característica que adoptan al sufrir la alteración de su membrana celular: eliptocitosis, esferocitosis, estomatocitosis, piropoiquilocitosis u ovalocitosis hereditaria, entre otras. Todas ellas son enfermedades raras y su prevalencia es  $< 1\%$  a nivel mundial (38). Una vez más vemos como las pruebas iniciales son muy útiles para sospechar la causa de la anemia y orientar su estudio siendo en este caso el frotis de sangre periférica es muy característico (**Figura 6**).

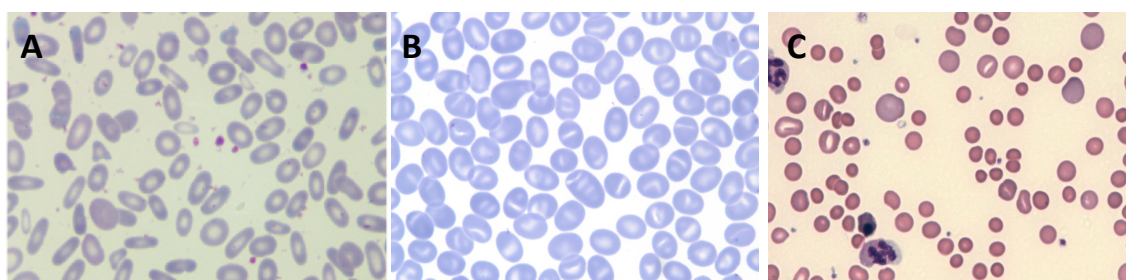


FIGURA 6. SEMIOLOGÍA ERITROIDE. A: ELIPTOCITOSIS. B: OVALOCITOSIS. C: ESFEROCITOSIS

#### 3.1.3.2.2 Anemias por déficit enzimáticos

Las anemias por déficit enzimáticos o eritroenzimopatías tienen una prevalencia aún menor que las membranopatías. Su fisiopatología comprende la pérdida de funciones de alguna de las vías metabólicas por las cuales el hematíe obtiene la energía necesaria

para desarrollar su cometido. La máxima representante de este grupo tan heterogéneo y raro es el déficit de Glucosa-6-fosfato-deshidrogenasa. En este caso, todas ellas van a presentar manifestaciones clínicas secundarias a la anemia, que está presente en todos los casos, aunque con una intensidad diferente, pudiendo ser reflejada como hemólisis crónica o aguda en el transcurso de la vida del paciente (4).

#### 3.1.3.2.3 Anemias por diseritropoyesis congénitas

Esta entidad es la anemia congénita menos frecuente y su incidencia, aunque se desconoce, se estima  $< 0,025\%$  de la población. Su fisiopatología radica en el trastorno de la maduración de la serie roja, produciendo alteraciones morfológicas y funcionales en el hematíe, consecuencia de ello se presentarán con mayor o menor anemia. La diseritropoyesis conduce a gran parte de los progenitores eritroides al “suicidio”, incluyendo a los reticulocitos, por tanto, nos encontraremos ante una anemia arregenerativa. Dependiendo del tipo de diseritropoyesis ante la cual nos encontremos, el mecanismo de producción de anemia variará entre la eritropoyesis ineficaz y la hemólisis periférica (39).

#### 3.1.3.2.4 Anemias microangiopáticas hereditarias

Corresponden a un grupo muy raro de enfermedades ya que suelen corresponder a la cara congénita de enfermedades que son adquiridas de manera habitual. En este caso los grandes representantes de estos cuadros son la PTT congénita y el Síndrome Hemolítico Urémico atípico (SHUa). En ambas entidades se describen alteraciones genéticas que predisponen al paciente en mayor o menor medida a tener una crisis

hemolítica que se acompañará de trombopenia en ambos casos. En el caso de la PTT congénita, la alteración genética predispondrá al paciente a sufrir una disminución de la función/producción de ADAMST13. En el caso del SHUa, la alteración genética se encuentra en genes del complemento que van a producir una desregulación en el mismo, llevando a una mayor destrucción de hematíes (4).

#### 3.1.3.2.5 Hemoglobinopatías

Sin duda este grupo de enfermedades es el gran representante de las anemias congénitas. Los resultados de los estudios epidemiológicos pueden llegar a ser muy diferentes, y encontramos publicaciones con alta fiabilidad que describen prevalencias de hasta del 20%. Sin embargo, cuando hablamos de alteraciones genéticas significativas estas cifras se colocan entre un 5-7% de la población mundial. Si nos movemos hasta nuestro país, la presencia de hemoglobinopatías en España se sitúa entre el 5-7% de la población, aunque los datos son antiguos y no hay duda de que el flujo migratorio hace que progresivamente esta prevalencia vaya en aumento (5,20).

Las hemoglobinopatías se dividen en 2 grandes grupos: talasemias y hemoglobinopatías estructurales. En España la mayoría de ellas están representadas por las alfa y beta talasemias, ya que las hemoglobinopatías estructurales tienen una incidencia estimada del 0,33% en España (40).

En las hemoglobinopatías se van a producir alteraciones en la producción de las cadenas de globinas, necesarias para unirse a los grupos hemos y formar la Hemoglobina. En el caso de las talasemias la alteración en las cadenas de globinas será cuantitativa mientras que, en el caso de las hemoglobinopatías estructurales, la alteración será cualitativa. Es

decir, si las alteraciones genéticas afectan a proteínas de síntesis de cadenas de globinas, nos encontraremos frente a una talasemia, mientras que, si las proteínas afectas forman parte de la estructura de la Hemoglobina, estaremos diagnosticando una hemoglobinopatía estructural. Además, son frecuentes las asociaciones entre ambas (talasemias + hemoglobinopatías) (41).

Debemos recordar que la Hemoglobina mayoritaria en el ser humano a partir de los 2 años de edad será la HbA1, conformada por un tetrámero 2 a 2 de cadenas de globina alfa y beta. Esta Hemoglobina constituirá aproximadamente el 95% de toda nuestra Hemoglobina durante el resto de nuestra vida. Las cadenas de globinas delta y épsilon tendrán su mayor papel durante la gestación del embrión y la vida temprana.

Las alteraciones genéticas en las talasemias pueden estar en las secuencias genéticas correspondientes a las cadenas alfa, cadenas beta, cadenas deltas, cadenas gamma, cadenas épsilon o en varias simultáneamente y su patrón de herencia es autosómico recesivo. Dependiendo de la cadena afectada, la talasemia adoptará el apellido correspondiente a la localización del fragmento genético alterado. Por ejemplo, en las beta-talasemias se verá disminuido el número de cadenas beta y, como ya hemos comentado, existen posibles combinaciones, pudiendo encontrarnos frente a una gamma-delta-beta-talasemias (4).

Si hablamos de hemoglobinopatías estructurales, la alteración producida dependerá de la localización del aminoácido mutado, produciendo a una anomalía en la función de la Hemoglobina. Lo más frecuente son mutaciones puntuales y su herencia es autosómica recesiva. Solamente las alteraciones genéticas que afecten a la solubilidad, estabilidad,

síntesis o función de la Hemoglobina serán clínicamente relevantes. La mayoría de ellas se denominan por las letras del alfabeto latino, pero según se ha seguido descubriendo nuevas variantes, han ido adoptando diferentes nombres, teniendo representantes a nivel nacional con la Hb Moncloa, por ejemplo (42).

Las más frecuentes son la Hemoglobina C, S y E, que además son las que mayor clínica pueden presentar. A nivel mundial la homocigosis para la Hb S es conocida como drepanocitosis (**Figura 3**) y es la mayor representante mundial de este grupo tan heterogéneo, probablemente por el efecto protector en estado heterocigoto frente a la malaria que conlleva (43).

La alta prevalencia actual y el aumento de incidencia en los países europeos y norteamericanos ha hecho que muchos países se planteen a las hemoglobinopatías como alerta sanitaria y hayan decidido mejorar su screening actual con el fin de aportar un correcto consejo genético a los futuros padres(44).

## 3.2 Introducción a la Inteligencia Artificial

### 3.2.1 Concepto e historia

El concepto de Inteligencia Artificial (IA) fue descrito por primera vez en 1950 en el artículo “Maquinaria computacional e inteligencia” publicado por Alan Turing, matemático de origen británico, donde describía la posibilidad de que las máquinas aprendieran a pensar como personas humanas. (45). Este mismo autor, en 1936, ya había descrito la computación como una serie de pasos ordenados y que podrían ser utilizados para resolver problemas (cálculos). Mucho tuvo que ver la segunda guerra mundial, y la necesidad de los británicos en descifrar la máquina “enigma” de los alemanes, lo que llevó a crear una máquina capaz de realizar cálculos con alta precisión y en un tiempo mucho menor que el ser humano<sup>1</sup>. Esta fue la primera vez que las máquinas aprendieron patrones de pensamiento humano y en una parte la superaron (en la velocidad). Aunque el concepto fuese descrito en 1950, el término IA fue nombrado por primera vez en 1959 por Arthur Samuel en su propuesta de algoritmo que aprendería a jugar a las damas (46), por lo que en esta década se asentaron los cimientos de la IA, cuya definición ha ido evolucionando a lo largo de los años.

Hoy en día encontramos múltiples definiciones de la misma, desde la definición de RAE que la define como la “disciplina científica que se ocupa de crear programas informáticos que ejecutan operaciones comparables a las que realiza la mente humana, como el aprendizaje o el razonamiento lógico” hasta algunas más simplificadas, como la del

---

<sup>1</sup> Esta historia fue llevada a la gran pantalla en 2014 con la película “The imitation game” en español “descifrando enigma”.

*Oxford English Dictionary* que la define como el “estudio y desarrollo de sistemas informáticos capaces de copiar el comportamiento humano inteligente”.

La IA durante las últimas dos décadas ha tenido un crecimiento notable, pero en los inicios tuvo dificultades de desarrollo que la mantuvieron detenida durante años (conocido como el invierno de la IA). Los mayores problemas estuvieron relacionados con el Hardware o la capacidad computacional disponible. Estos problemas no impidieron que técnicas de *Machine Learning* tradicional aparecieran en la década de los 80 y 90, como los árboles de decisión o las máquinas de vector de soporte, pero sí que impidió la evolución de las redes neuronales, que tendrían que esperar a una mejora en los recursos computacionales de la época.

Ya en el siglo XXI, avances tecnológicos como microprocesadores de gran capacidad y la informatización de los sistemas (lo que produjo la disponibilidad de datos que conocemos como la época del Big data), junto con la capacidad de sacar provecho de esos datos (minería de datos), han permitido retomar la carrera de la IA, esta vez centrada en la proliferación de redes neuronales.

Los sistemas de IA que existen en la actualidad están basados en el aprendizaje de labores concretas, aprendidas en su mayoría a base de ejemplos y experiencias aportadas por el ser humano. El objetivo final está basado en que los sistemas aprendan patrones característicos que permitan separar diferentes clases o grupos. Esto limita la experiencia de dichos sistemas a los ejemplos en los que han sido entrenados. Por ejemplo, podemos enseñar a un algoritmo a diferenciar entre coches y camiones, pero si le suministramos una imagen de una moto intentará clasificarlo en alguna de esas dos categorías, un coche o un camión, ya que en su universo el concepto de moto no existe.

Este concepto es importante entenderlo para entender la mayoría de errores que se pueden dar en IA.

De forma parecida a los humanos, estos sistemas aprenden de ejemplos. Sólo que los sistemas de aprendizaje de nuestros algoritmos inteligentes son mucho más ineficientes y necesitan de muchísimos ejemplos para empezar a funcionar. En líneas generales, cuanto mayor experiencia le aportemos (datos), mayores serán las probabilidades de encontrar las características que identifiquen las muestras y mejores resultados obtendremos. Estos resultados son validados mediante el uso de métricas que miden la precisión, especificidad y la sensibilidad de los algoritmos(47).

Las muestras que se utilizan en el proceso de entrenamiento de un algoritmo inteligente juegan un papel determinante en la calidad del mismo. En este sentido, es importante conocer el Big Data, la Minería de datos y el Internet de las cosas. La aparición de los *Smartphones* ha jugado un papel muy importante en la obtención de grandes cantidades de datos, pero existen muchas otras fuentes de datos que pueden pasarnos desapercibidas, como el manejo de las cookies, las plataformas de películas o de música, los relojes inteligentes, televisores, las cámaras de cualquier dispositivo, etc.

Pese a la velocidad a la que evoluciona la Inteligencia Artificial en todo el mundo, las máquinas inteligentes a día de hoy se basan en lo que se conoce como Inteligencia Artificial débil o *Artificial Narrow Intelligence*. Es decir, son sistemas informáticos creadas para realizar funciones concretas y solo esas. Por tanto, la idea de sustituir al ser humano por sistemas de IA son ciencia ficción en el año 2023 (48).

Sin embargo, estos sistemas de IA sí pueden apoyar muchas profesiones, mejorando resultados y aumentando la eficacia en múltiples ámbitos, sobre todo en aquellos

repetitivos y sistemáticos. Por ejemplo, en 2021, *DeepMind*®, un sistema de IA creado por Google, demostró ser capaz de predecir la estructura tridimensional de una proteína en segundos, hecho que normalmente ocupaba meses o años en el trabajo de investigación de nuevas moléculas y tratamientos. Los resultados se han compartido con toda la comunidad científica y según la revista *Science* fue el avance científico más importante de su año por la mejora en desarrollo de medicinas(49,50).

La Inteligencia Artificial está destinada a formar parte de nuestra vida, probablemente produciendo mejoras como fueron la aparición del teléfono móvil o el teléfono inteligente. Las grandes empresas internacionales como Google, Microsoft, Facebook, así como las grandes naciones trabajan en ser las más actualizadas en este campo, por sus potenciales usos. Los debates que plantea el uso de la IA van más allá de lo técnico. Los cambios a nivel sociocultural y económico en los que el uso de IA puede desembocar, van a dar forma a nuestro mundo en los próximos años. Especialmente en el ámbito sanitario, es importante que sepamos utilizarla para apoyarnos en ella para mejorar la eficiencia de nuestro trabajo, teniendo claro que la relación médico-paciente seguirá siendo indispensable para conseguir la combinación perfecta entre la medicina personalizada y medicina de alta precisión.

### 3.2.2 Clasificación y breve introducción a los métodos de IA

La mayoría de la Inteligencia Artificial en la que se trabaja a día de hoy está basada en el Aprendizaje Automático de Maquinas o *Machine Learning* (ML) y en el Aprendizaje Profundo o Deep Learning (DL), aunque esta segunda la podemos considerar una rama de la primera (Figura 7).

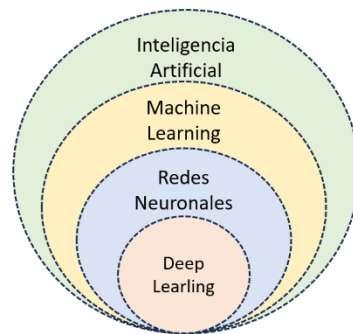


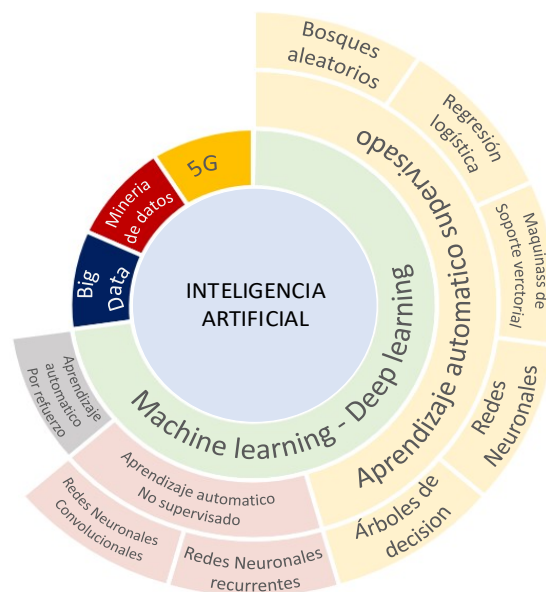
FIGURA 7. CLASIFICACIÓN GENERAL DE LA INTELIGENCIA ARTIFICIAL

Dentro de cada uno de estos sistemas encontremos ramas más específicas, como el procesamiento de señales (*signal processing*) o el procesamiento del lenguaje natural (*natural language processing*), tan de moda en los últimos años con sistemas tan normalizados entre nosotros como el de *Alexa* de Amazon® o los dispositivos asistentes personales de Google® (51). Simplificando mucho el concepto, realmente el procesamiento del lenguaje natural es cualquier forma de procesamiento de textos.

El ML nació con el objetivo de analizar grandes cantidades de datos de forma automática. Uno de los ejemplos más antiguos es el de la bandeja *Spam* de correos, donde el algoritmo, según una serie de patrones repetidos en todos los correos de este tipo, es capaz de seleccionar con una alta precisión qué correos no nos interesan o pertenecen a anuncios automáticos, para clasificarlos como *Spam*. Este funcionamiento está basado en el **Aprendizaje automático supervisado**. Continuando con este ejemplo,

este tipo de aprendizaje está basado en mostrar al sistema una alta cantidad de correos que han sido previamente etiquetados como *Spam vs No Spam* para que la maquina reconozca las características que posteriormente le ayudarán a ser capaz de clasificar los correos por él mismo (52).

Si por el contrario hablamos de **métodos de aprendizaje automático no supervisados** le mostraremos a la máquina un conjunto de datos sin ningún tipo de etiqueta elegida por nuestra parte. El algoritmo analizará las características de estos datos y creará grupos de datos con características similares entre sí. Este tipo de aprendizaje necesita un volumen de datos mayor que el aprendizaje supervisado, ya que nuestra participación es mucho menor y no participamos en la calibración del algoritmo (**Figura 8**).



**FIGURA 8. CLASIFICACIÓN DE LOS DIFERENTES MODELOS DE INTELIGENCIA ARTIFICIAL QUE SE TRATAN EN ESTA TESIS DOCTORAL.**

Para finalizar, también existe el **aprendizaje automático de máquinas por refuerzo**. En este caso cada vez que la maquina etiquete un dato, le daremos una recompensa si acierta o le “castigaremos” si se equivoca, corrigiéndole y enseñándole como solucionar

ese problema. Siguiendo con el ejemplo de los correos, cada vez que recibimos un correo a la bandeja de entrada y pensamos que debe estar en SPAM, lo redirigimos a esa carpeta. Lo que realmente estamos haciendo es corrigiendo a la maquina en que ese correo lo ha clasificado mal. Podemos decir que los algoritmos aprenden como los niños: con los ejemplos y el refuerzo. Incluso podemos pensar que estos programas, en parte, se codifican a sí mismos, basados en nuestros ejemplos (14).

Como comentamos previamente, el ML y sus métodos de clasificación existen desde hace 50 años, pero al no existir los recursos computacionales necesarios para su correcto funcionamiento, estos sistemas perdieron su atractivo. Durante estos años se desarrollaron los métodos tradicionales de menor complejidad matemática. Entre los métodos tradicionales más relevantes conocemos la Regresión Logística, los Árboles de Decisión (DT, del inglés *Decisión Trees*), el Bosque Aleatorio (RF, del inglés *Random Forest*) y Maquinas de Vector Soporte (*SVM*, del inglés *Support Vector Machine*). **Figura 8**

Los avances tecnológicos han permitido el desarrollo de las redes neuronales artificiales, basadas en la idea de imitar el sistema nervioso humano: tras recibir una información y tras un número variable de conexiones entre los datos recibidos, obtendremos un respuesta o resultado (51).

### 3.2.2.1 Regresión lineal

Es el método más básico de aprendizaje automático supervisado, pero sigue teniendo cierta utilidad y es necesario de conocer para entender los siguientes. El objetivo de la regresión lineal es encontrar la función lineal que mejor se ajuste a los datos, es decir, la función que minimiza las diferencias entre los valores reales y los valores predichos

por el modelo. Un ejemplo validado es la relación que existe entre el tabaco y el cáncer de pulmón, donde mediante el Índice de paquetes/año (IPA) podremos predecir la aparición de cáncer de (Figura 9). Aunque existe una clara asociación, existen pacientes grandes fumadores que no padecerán cáncer y viceversa. Sin embargo, estos pacientes no serían clasificados correctamente por el algoritmo, al basarse únicamente en 1 variable (9).

### 3.2.2.2 Regresión Logística

La regresión logística es otro modelo de aprendizaje supervisado utilizado en IA para predecir la probabilidad de que ocurra un evento binario. A diferencia de la regresión lineal, la regresión logística utiliza una función logística (función matemática sigmoidea en forma de S) para modelar la relación. Siguiendo con el ejemplo, la regresión logística nos permite clasificar de mejor manera a un mayor número de pacientes enfermos de cáncer de pulmón partiendo de los datos del IPA (Figura 9).

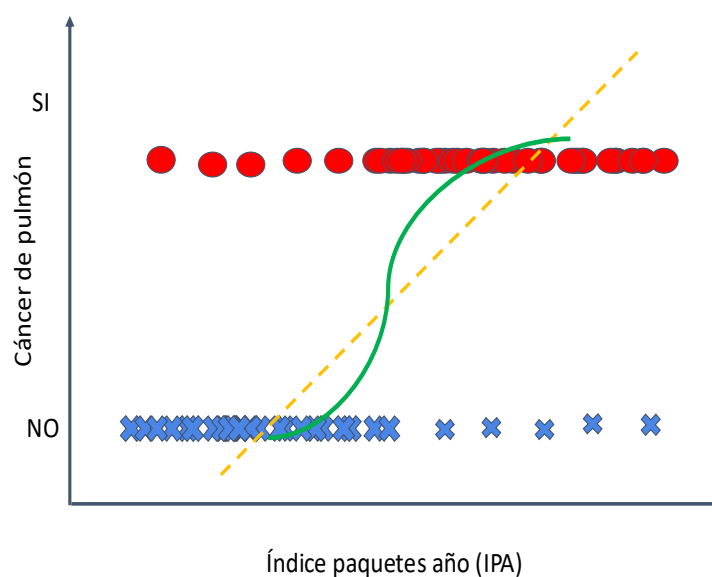


FIGURA 9. REPRESENTACIÓN DE LA REGRESIÓN LINEAL (LÍNEA AMARILLA) Y REGRESIÓN LOGÍSTICA (LÍNEA VERDE) UTILIZANDO UN MODELO DE ASOCIACIÓN VALIDADO (TABACO Y CA. PULMÓN).

### 3.2.2.3 Máquinas de Soporte Vectorial

La Máquina de Soporte Vectorial es un modelo de aprendizaje supervisado utilizado en IA para *clasificación* y *regresión*. Con SVM se pueden crear algoritmos con más de 2 dimensiones (hiperplanos), creando una dimensión por cada variable introducida (**Figura 10**) (48,49).

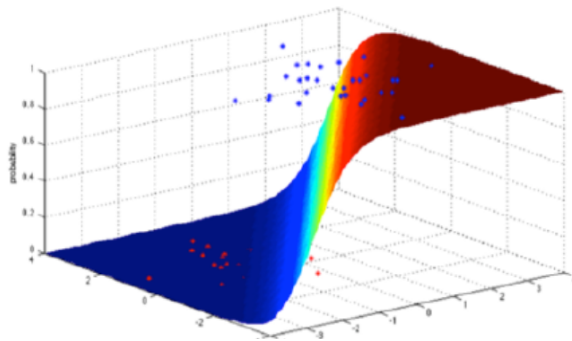


FIGURA 10. EJEMPLO DE HIPERPLANO DE 2 DIMENSIONES

En el caso de la *clasificación*, el objetivo de una SVM es encontrar el hiperplano que mejor separa las clases de datos en un espacio de alta dimensión. Un hiperplano es una superficie que divide el espacio en dos regiones. En el caso de la SVM, se busca el hiperplano que maximiza la distancia entre las dos clases de datos más cercanas (**Figura 11**).

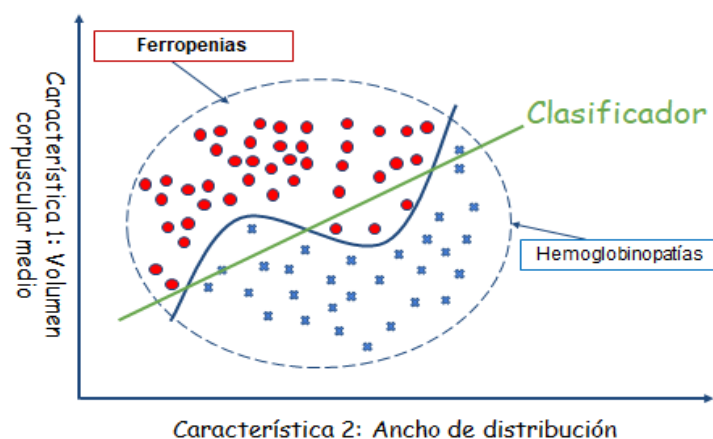


FIGURA 11. REPRESENTACIÓN DEL FUNCIONAMIENTO DE LA MÁQUINA DE SOPORTE VECTORIAL

En el caso de la *regresión*, la SVM se utiliza para encontrar una función que se ajuste a los datos de entrenamiento y tenga el menor error posible entre los valores predichos y

los valores reales. La **Figura 11** representa este vector clasificador, que debemos imaginar con más de 2 dimensiones para acercarnos más a la realidad, algo parecido a la **Figura 10**. En nuestro ejemplo, representado únicamente por 2 variables o características, vemos como se separan 2 grupos de anemias.

Las SVM utilizan la función de *kernel* para transformar los datos en un espacio de alta dimensión en el que sea más fácil encontrar un hiperplano de separación (**Figura 12**).

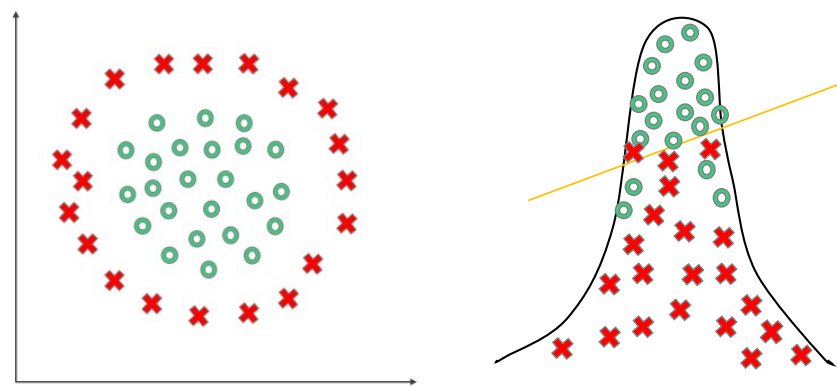
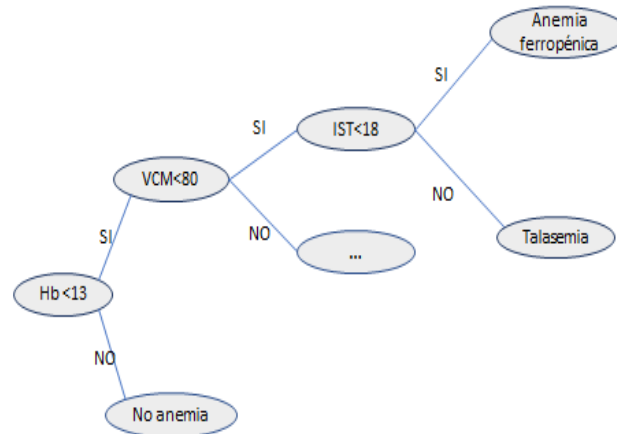


FIGURA 12. IMAGEN REPRESENTATIVA DE LA FUNCIÓN DE KERNEL (LÍNEA AMARILLA).

### 3.2.2.4 Árboles de decisión

Este modelo de aprendizaje automático supervisado está basado en estructuras similares a un diagrama de flujo que se va dividiendo en función a la respuesta a una serie de preguntas que se le formulan al sistema (Figura 13). Estos modelos están



**FIGURA 13. EJEMPLO BÁSICO DE ÁRBOL DE DECISIÓN PARA EL DIAGNÓSTICO DE ANEMIAS**

compuestos de nodos o preguntas y ramas o respuestas. Cada nodo representa una pregunta sobre una característica específica de los datos de entrada. Las respuestas a estas preguntas llevan a la siguiente pregunta y así sucesivamente, hasta llegar a una hoja del árbol que representa una clasificación o predicción final (55). Dos de los métodos más utilizados, y los utilizados en esta tesis son:

a) Bosques aleatorios o Random Forest (RF)

Es el más clásico de este sistema de clasificación. Lo ideal es encontrar nodos que dividan al grupo en 2 grupos similares (en tamaño y características). En un RF se construyen múltiples árboles de decisión utilizando diferentes subconjuntos de datos de entrenamiento y características seleccionadas mediante las preguntas elegidas (**Figura 14**). Posteriormente, se elige el árbol con sus correspondientes nodos que mejor clasifique el problema original (56,57).

b) XGBoost (*eXtreme Gradient Boosting*):

Es un algoritmo de aprendizaje automático que utiliza múltiples árboles de decisión para realizar predicciones. La idea detrás de XGBoost es entrenar un conjunto de árboles de decisión de forma secuencial, donde cada árbol intenta mejorar las predicciones del modelo anterior. La gran diferencia con RF es el trabajo secuencial, es decir, el sistema crea un árbol de decisión y, sobre ese mismo árbol, va creando nuevas opciones de nodo y se reevalúa constantemente (**Figura 14**). Si ha mejorado la capacidad de clasificación, deshecha el árbol anterior y sigue trabajando de la misma manera con el nuevo árbol. Es un sistema muy utilizado en la actualidad (55).

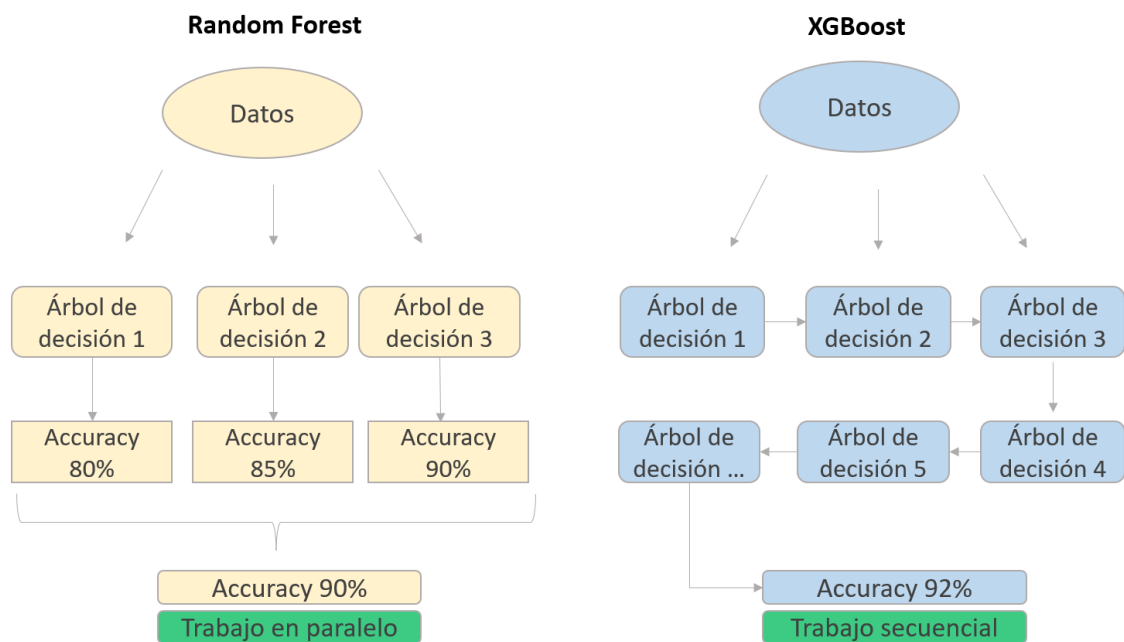
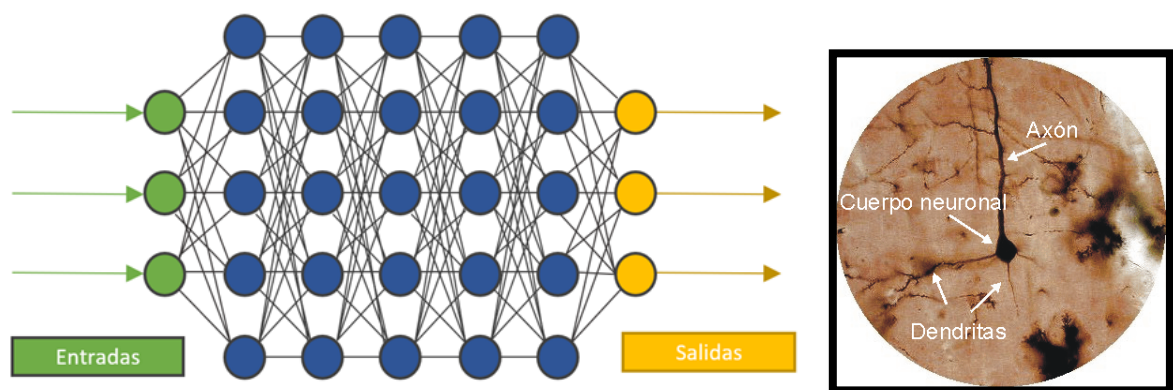


FIGURA 14. REPRESENTACIÓN ESQUEMÁTICA DE DOS DIFERENTES FORMAS DE CREAR UN ALGORITMO BASADO EN ARBOLES DE DECISIÓN. DIFERENCIAS ENTRE LOS 2 MÉTODOS UTILIZADOS EN ESTA TESIS (RANDOM FOREST Y XGBoost).

### 3.2.2.5 Redes neuronales artificiales

Las redes neuronales artificiales tratan de asemejarse a las neuronas humanas: millones de interconexiones (dendritas) que van de nodo a nodo (neuronas), donde se procesa la información y desde donde salen más conexiones hacia las siguientes unidades nodales, creando una potente y rápida red de comunicación y procesamiento (**Figura 15**).

Cada unidad neuronal artificial se encarga de recibir información y procesarla mediante una función matemática. Cuando esta neurona ha procesado la información, vuelve a transmitir la información a la siguiente neurona, donde habrá otro nodo con una función matemática diferente, hasta llegar a la última neurona. Cada uno de estos niveles de neuronas serán denominadas capas neuronales. A este respecto, el término Deep Learning, corresponde a una Red Neuronal creada con varias capas, que pueden considerarse 3 o más según diferentes autores, lo cual es un debate abierto (9,14,55). Las neuronas o nodos de la primera capa reciben diferentes estímulos denominados *INPUTS* (información de entrada). Las neuronas entrelazadas en diferentes capas van transmitiéndose la información y aplicando las funciones matemáticas que acompañan a cada neurona (elegidas por el arquitecto ingeniero informático) para dar un resultado final (*OUTPUT*).



**FIGURA 15. EJEMPLO DE RED NEURONAL ARTIFICIAL, COMPARADA CON UNA NEURONA HUMANA.**

Hoy en día se llegan a crear redes neuronales de hasta 100 capas de profundidad, de donde viene la idea del aprendizaje profundo. A mayor número de capas, se necesitarán mayores cantidades de datos para tener suficientes ejemplos que representen todas las opciones de clasificación de una manera adecuada. A su vez, una mayor cantidad de datos conllevan una menor posibilidad de supervisión humana, ya que no somos capaces de gestionar el *Big data* de manera manual. Es por ello que la minería de datos y el *Big data* juegan un papel muy importante en este sistema. Así mismo, la participación del ingeniero informático, el arquitecto del algoritmo, es mayor que en los métodos tradicionales de ML, ya que tienen que ajustar el algoritmo hasta conseguir el mejor resultado posible de clasificación.

### 3.2.3 Aplicaciones de la Inteligencia Artificial en el mundo real

En el año 2023 vivimos rodeados de sistemas informáticos equipados con Inteligencia Artificial. Como ya hemos dicho, la IA se alimenta de datos y para recoger esos datos necesitamos dispositivos conectados a la red. A fecha de hoy se estima que lleguemos a tener en torno a los 30 millones de dispositivos conectados a la red (58).

Uno de los proyectos con más repercusión social e inversión económica acapara de IA a día de hoy, es el del coche autónomo. Los resultados obtenidos en cuanto a seguridad son muy robustos (objetivo primario) y en Phoenix (EEUU) ya existen taxis con capacidad autónoma que se encuentran en circulación desde 2017 y se prevé que lleguen al resto del mercado mundial en 2030 (7,59).

Pero esto es solo la punta del iceberg ya que existen un sinnúmero de procesos de IA con los que convivimos en el día a día: Los sistemas de búsqueda de Google®, las listas de reproducción propuestas por Spotify®, las cuentas recomendadas en Instagram® y las decenas de anuncios que recibimos mientras surfamos por internet están basados en algoritmos inteligentes. De la misma manera, el uso de nuestros relojes inteligentes o los sistemas de reconocimiento facial que utilizamos para desbloquear nuestro *Smartphone* también han evolucionado notablemente desde sus inicios en 1991, así como otro gran número de aplicaciones que no son el objetivo de esta tesis (60,61).

Mención aparte merece comentar que en los últimos meses (octubre del año 2022), hemos asistido a la aparición de una nueva aplicación de IA que ha revolucionado las redes sociales y el mundo científico: *ChatGPT* o *ChatGPT4* en su última versión (62). Este sistema de IA se creó con el objetivo de producir escritos semántica y gramaticalmente

correctos. Para ellos se utilizó una base de datos con millones de artículos, poemas, novelas, noticias y otros documentos de múltiple índole. La máquina ha conseguido aprender a mantener una conversación con un humano, siendo prácticamente indistinguible de otro ser humano. Además, al aprender el lenguaje a partir de una serie de documentos, ha retenido los conocimientos descritos en dichos documentos y finalmente se ha llegado a utilizar como un consultor, aunque no fuese su objetivo principal.

OpenAI®, la empresa diseñadora del sistema, ha iniciado relaciones con Microsoft y a día de hoy está implementado en su buscador *Bing*® y en sus servidores *Azure*®(63,64). Este sistema nuevo de ChatBot promete ser muy trascendente en los próximos años, incluido a nivel científico, interesándose por él revistas internacionales de prestigio como *Nature* y mostrando cual puede ser su potencial uso en el futuro, entre las que se encuentran aplicaciones en documentación sanitaria, diagnóstico, investigación y educación (65,66).

### 3.2.4 Aplicaciones de la Inteligencia Artificial en salud

Aunque el uso de la Inteligencia Artificial dentro del sistema de salud pública nacional es limitado a día de hoy, la IA se está introduciendo en ensayos y estudios con el objetivo de validar su utilidad y formar parte de medicina del siglo XXI, tal y como se introdujo la informática en medicina hace ya años (67) . Claro ejemplo de ello es el compromiso que están adquiriendo algunas revistas como el *New England Journal of Medicine*, que ya ha publicado que en el año 2024 inaugurarán una revista únicamente para artículos de IA en medicina y que llamarán NEJM-IA (68). En este contexto también se están definiendo cuáles serán las reglas de juego a la hora de considerar y evaluar las nuevas publicaciones (69,70), así como cuáles serán las normas básicas a seguir para llevar a cabo ensayos clínicos asociados a IA con las guías CONSORT-IA (71), donde la transparencia de datos y métodos cada vez acapara mayor interés (72), por su dificultad para conseguirla cuando hablamos de IA.

De manera global, se prevé que la IA impacte en tres grandes campos de la sanidad. El primero es en el trabajo del médico, ya que podrán disminuir la carga de trabajo, manteniendo la precisión de diagnóstico para poder dedicar mayor tiempo en la atención del paciente (73,74). El segundo campo de mejora es el de los sistemas sanitarios, mejorando el flujo de trabajo, disminuyendo las listas de esperas y disminuyendo las interconsultas entre distintos especialistas. Todo ello gracias a sistemas inteligentes que aseguren una correcta interpretación de imágenes o datos analíticos, así como apoyando el proceso de análisis de coste-eficacia con los que mejorar las pruebas de screening de enfermedades realizadas a nivel nacional (75). El tercer nivel de impacto, y probablemente el más importante, es resultado de los 2

puntos previos y es la mejora en el diagnóstico, screening y tratamiento de enfermedades, resultando en una mejora en la atención del paciente (76–78). En este sentido cobra gran interés la posibilidad de ofrecer esta mejora de manera equidistante en todos los puntos del mundo, siendo en aquellas zonas menos favorecidas o peor conectadas (como las zonas rurales), en donde encontrar un sanitario especialista se convierte en una tarea imposible (79,80). No debemos olvidar el papel de la adaptación de la robótica en los campos quirúrgicos, algo que ya vemos en el día a día de nuestros hospitales con el robot *Da Vinci*, el cual permite eliminar el temblor fisiológico humano y ampliar hasta 10 veces el campo quirúrgico. Se han publicado en Dinamarca sistemas robotizados para el procesamiento analítico capaces de llevar a cabo el análisis de hasta 3.000 muestras diarias, llevando a cabo el proceso tras la obtención de la analítica (81,82).

Aunque hay muchos retos que afrontar, podemos afirmar que la IA se ha introducido en medicina en numerosos campos, casi todos orientados al apoyo en el diagnóstico de enfermedades, como Alzheimer, Cáncer, Diabetes e Hipertensión entre muchas otras (83). La identificación y clasificación de imágenes, el reconocimiento de características de pacientes mediante el uso de lenguaje natural o la detección de enfermedades con sensores digitales son algunas de las técnicas utilizadas que podemos ver. De hecho, el reconocimiento celular en las fórmulas leucocitarias está basado en aplicaciones inteligentes integradas en nuestros analizadores del laboratorio desde hace años, dándonos señales de alarma cuando es preciso para que se revisen las analíticas por especialista (hematólogo o analista).

Existen aplicaciones inteligentes de todo tipo utilizadas en el campo sanitario, incluso algunas que llaman la atención por su originalidad. Entre ellas encontramos el diagnóstico de enfermedades mediante el reconocimiento de las características faciales en enfermedades como el autismo o el Síndrome de Turner (84,85); el reconocimiento de fases maníacas en el paciente con trastorno bipolar a partir de sus publicaciones en redes sociales (86), o aplicaciones capaces de predecir el delirium postoperatorio a partir de datos analíticos como la Hemoglobina, la función renal y la urea en sangre (87). Pese a la innovación que representan estas aplicaciones, la metodología utilizada debe mejorar en algunos de estos trabajos para obtener mayor validez científica.

Si volvemos hacia atrás y valoramos los avances que han supuesto las redes neuronales en el contexto de la IA en salud, tenemos que hablar del reconocimiento de imágenes. Previamente a su aparición, los sistemas de aprendizaje automático tradicional (Support Vector Machine, Random Forest, etc.) habían ya conseguido buenos resultados en el análisis de datos, pero no cuando hablábamos de reconocimiento de imagen. Las redes neuronales han sido la gran disrupción en este contexto y podemos ver como se han ido desarrollando en prácticamente la totalidad de las especialidades.

Por ejemplo, en cardiología se han diseñado desde algoritmos diagnóstico de arritmias basados en las imágenes de los electrocardiogramas publicados en 2016 hasta modernos algoritmos que son capaces de mejorar la estratificación de las patologías cardiovasculares publicados en 2018 (88,89).

En oftalmología hay múltiples publicaciones donde se han utilizado imágenes de fondo de ojo para diagnosticar retinopatía diabética consiguiendo un AUC de las curvas ROC

del 0,97 (90), así como otras publicaciones que tratan de determinar el sexo de una persona mediante imágenes similares(91).

Existen especialidades donde la imagen o la valoración óptica es imprescindible, como la dermatología(92) o la anatomía patológica(93), donde existen publicaciones que demuestran que la participación de algoritmos de Inteligencia Artificial apoyando al diagnóstico a un patólogo especialista pueden mejorar los resultados de diagnóstico (94,95).

Sin embargo, el mundo de la radiología es el más incurrido por la IA, lo que ha llevado a algunos individuos a equivocarse en el enfoque del tándem IA-Radiología como un mundo donde compiten hombres contra máquinas(96). Desde el punto de vista científico, hace años que se publican artículos donde se describen mejoras en el diagnóstico cuando existen colaboraciones IA-medicina en radiología. Un ejemplo de vida real es el apoyo que ofrecen estos sistemas en el screening de cáncer de mama, donde ha sido publicado por *McKinney et al.* un sistema de IA que consigue un Área bajo la curva (AUC-ROC) de 0,889%, mejorando hasta en 11,5% los resultados previos (97). En este contexto, la pandemia por COVID también se ha visto incurrida por la IA en radiología con sensibilidades y especificidades en torno al 90% a la hora del reconocimiento de neumonías virales (98).

No obstante, la necesidad de grandes cantidades de datos aún limita el procesamiento y aprendizaje, que se acentúa cuando se utilizan ciertas técnicas como la resonancia magnética o la tomografía computarizada. Teniendo en cuenta la gran complejidad de estos estudios (en comparación con radiografías simples o mamografías), la cantidad de datos necesaria para conseguir un aprendizaje correcto de todas las variables y posibles

diagnósticos es considerablemente mayor, con lo que supone el mayor factor limitante en la radiología (99).

### 3.3 Aplicaciones de Inteligencia Artificial en hematología

Dado que una parte importante de la hematología comprende el diagnóstico de enfermedades, la Inteligencia Artificial puede llegar a tener un grado de participación importante. Como muestra de ello, en los últimos años hemos visto aplicaciones de IA en el análisis de datos analíticos, valoración de frotis de sangre periférica y médula ósea, citometría de flujo, estudios genéticos y predicciones de riesgo, como el de tener un síndrome de injerto contra huésped tras un trasplante medular, con resultados de todo tipo, pero obteniéndose en la mayoría de ellos Accuracy mayores al 80% (100–102). La mayoría de estas aplicaciones están basadas en el análisis de imágenes (celulares), aunque encontramos también algoritmos aplicados al análisis de datos tabulados. Generalmente los trabajos en los que se implica la IA están divididos entre las aplicaciones dedicadas a la citología, la citogenética, la biología molecular y la citometría de flujo, donde profundizaremos más adelante (103–106).

Además del diagnóstico, también se están explorando otros campos, como el de la valoración y respuesta a los tratamientos o la creación de nuevos índices pronósticos en oncohematología, donde encontramos una participación importante del Grupo español de Mieloma múltiple, que está trabajando con la IA para conseguir nuevos índices pronósticos en Mieloma múltiple(107,108).

Pese a las elevadísimas publicaciones en las que se menciona la medicina personalizada mediante el apoyo de la IA en los laboratorios, a día de hoy seguimos notando la ausencia o escasez de aplicaciones de IA en el sistema público de salud (13,109,110).

*CellaVision*<sup>®</sup>, un sistema en el que una máquina clasifica las diferentes poblaciones celulares de un frotis de sangre periférica, y que posteriormente valida un hematólogo,

es el sistema de IA más utilizado en hematología (111,112), aunque no el único existente con autorización de la FDA para el reconocimiento de celularidad en sangre periférica, como es *Scorpio*<sup>®</sup>(113). Dada la escasa presencia a día de hoy de las aplicaciones de IA instauradas en el hospital o centro de salud, probablemente con esta única aplicación incluida de manera rutinaria en los laboratorios, la hematología ya se coloque por delante de la mayoría de especialidades en la carrera de la IA aplicada a la sanidad (114–116).

La mayoría de los algoritmos dedicados al análisis de datos tabulados están basados en sistemas de *Machine Learning* tradicional (es decir sin el uso de redes neuronales). Pese a que no se aplicaban redes neuronales, desde 2017 se han publicado resultados con precisiones en torno al 85% para el diagnóstico de enfermedades hematológicas. Eso sí, con la necesidad de inmensas cantidades de datos, donde se incluyen datos de bioquímicas además de los del propio hemograma (117).

Como ya hemos comentado previamente, la progresiva y exponencial mejora de los sistemas informáticos en los últimos años han permitido que las redes neuronales pudieran desarrollarse sin limitantes y con ello mejorar el reconocimiento de imágenes. Es por ello que prácticamente casi todo lo publicado antes del 2019, estará enfocado en sistemas de *Machine Learning* tradicional y desde el 2019 en adelante con alta probabilidad harán referencia a Redes neuronales y a su aplicación al reconocimiento de imágenes, tanto en hematología como en otras especialidades(118–124).

Si echamos la vista atrás y revisamos las publicaciones de hace 15 años, vemos como estaban centradas en describir cuales eran las posibilidades que podría ofrecer la IA en la hematología (100) , sin embargo las últimas publicaciones ya nos describen cuales son

los campos en los que se han encontrado y las grandes promesas de mejora que se esperan de esta relación (125).

A continuación, explicaremos el estado del arte en cada uno de los campos de la hematología.

### 3.3.1 Inteligencia Artificial en hematología mediante el análisis de datos analíticos

Cuando nos referimos al análisis y procesamiento de datos tabulados tenemos que saber que los resultados conseguidos con los sistemas tradicionales de Machine Learning (árboles de decisión y máquinas de soporte vectorial en su mayoría) han conseguido precisiones superiores al 80% durante los últimos años (117) . Probablemente por ello, las redes neuronales no se han preocupado en mejorar estos sistemas, teniendo como objetivo principal el análisis de imagen (126,127).

Aunque encontramos trabajos en los que se clasifican enfermedades oncohematológicas mediante resultados analíticos, la mayoría de estos trabajos están centrados en el diagnóstico de anemias, a lo que le dedicaremos un apartado específico. No obstante los resultados encontrados consiguieron en algunos escenarios concretos precisiones de hasta el 84,9% y un área bajo la curva (AUC) de 93,5% (128,129) .

### 3.3.2 Inteligencia Artificial en la clasificación de imágenes celulares

La valoración citológica mediante el microscopio óptico ha sido durante años y sigue siendo a día de hoy la piedra angular en los diagnósticos hematológicos, que poco a poco se han ido apoyando en la citometría de flujo, la citogenética y la biología molecular.

Aunque el *CellaVision*<sup>®</sup> es de los únicos sistemas incluidos en los laboratorios del sistema sanitario, cuando revisamos las publicaciones de IA en hematología, vemos cómo existe en los últimos años una revolución en la búsqueda del mejor algoritmo de clasificación celular (73,74,83,95,96).

Las aplicaciones publicadas son de todo tipo, desde las más básicas publicadas hace ya 20 años (2003) en las que se realiza la clasificación de enfermedades hematológicas a partir de los citogramas, pasando por publicaciones más recientes (2014) en las que se clasifican glóbulos blancos vs glóbulos rojos. Artículos más recientes tratan de diagnosticar hasta 7 enfermedades hematológicas mediante la clasificación de hasta 19 clases celulares en médula ósea, donde la clasificación celular coincide en más del 60% entre hematólogos y máquinas, siendo el reconocimiento de los linfocitos donde peores resultados se obtienen (100,131–133)

De hecho ya se han publicado aplicaciones móviles que nos permiten diferenciar 5 patologías hematológicas a partir de las imágenes de los hematíes con una precisión cercana al 100% y ya existen analizadores novedosos en los que se utiliza Inteligencia Artificial para el reconocimiento rutinario de células en el frotis de sangre periférica con una sensibilidad y especificidad mayor al 95% (134–136).

Reciente se han publicado algoritmos capaces de diferenciar 21 tipos diferentes de células en la medula ósea con resultados muy prometedores, llegando a clasificar correctamente al 84% de la celularidad tras una validación externa, punto muy interesante a la hora de utilizar el algoritmo en otros centros (130,137).

La utilización de redes neuronales convolucionales (redes formadas por un gran número de capas de neuronas o nodos) permiten incluso el reconocimiento celular a partir únicamente de un fragmento de la imagen celular (debido a superposiciones de la celularidad o fin de la extensión), lo que aporta mejores resultados de precisión (138).

Nuevos sistemas de IA basados en redes neuronales convolucionales, con grandes cantidades de datos, combinan redes neuronales (utilizadas para extraer características celulares morfológicas) y Maquinas de Soporte Vectorial (con las que se enseña a la maquina las características típicas de las células asociadas a una enfermedad). Los resultados obtenidos llegan a alcanzar en algunos experimentos precisiones de más del 97% en ciertas enfermedades hematológicas (139).

### 3.3.3 Inteligencia Artificial en la citometría de flujo

La citometría de flujo es un campo menos trabajado por la Inteligencia Artificial que el reconocimiento de imágenes celulares. Aunque pueda parecerlo que la citometría de flujo es altamente objetiva, la experiencia y el conocimiento del hematólogo analizador sigue teniendo el mayor peso de la balanza para lograr un diagnóstico correcto. La variabilidad interobservador es la traducción de los errores humanos que se llegan a cometer. Con intención de evitarlos, se está trabajando en aplicaciones de redes neuronales, tanto para el procesamiento de imágenes previamente trabajadas por un hematólogo, como en imágenes “desnudas” (106).

La Sociedad Internacional para el Avance de la Citometría, dentro de su proyecto *FlowCat*, publicó en 2020 y 2021 los resultados de un estudio realizado con más de 20.000 muestras donde trataron de clasificar 8 tipos de neoplasia linfoproliferativas. En el 70% de las muestras consiguieron clasificar bien las enfermedades con un 95% de confianza, pero en las enfermedades con menos datos disponibles (linfoma de la zona marginal, leucemia prolinfocítica y linfoma linfoplasmocítico), la precisión bajaba hasta el 61,61 y 64% respectivamente, reflejando una vez más la necesidad de grandes cantidades de datos. Para ello utilizaron sistemas de redes neuronales convolucionales basadas en aprendizaje automático supervisado. En resumen, el objetivo de *FlowCat* es clasificar automáticamente los datos de la citometría de flujo en diagnóstico de linfomas sin necesidad de realizar ninguna selección manual inicial. El objetivo futuro está centrado en aumentar el número de muestras y disminuir el número de paneles utilizados (140,141).

Por otro lado, se está estudiando el valor que aportaría la participación de la IA en la enfermedad mínima residual (EMR) mediante la automatización de la citometría de flujo, es decir, para valorar la evolución y respuesta del paciente al tratamiento. Una vez más, el grupo español de mieloma múltiple (dentro del consorcio *Euroflow*) tiene una gran participación y los resultados publicados son muy alentadores, donde se observa una concordancia de datos (entre el algoritmo y los estudios considerados *Gold Standard*) que supera el 95% de los casos estudiados(142,143).

### 3.3.4 Inteligencia Artificial en aplicaciones clínicas de hematología

Las aplicaciones de IA en la práctica clínica general son muy numerosas pero la mayoría ellas se centran en la predicción de riesgos: predicción de osteoporosis, necesidad de cuidados intensivos durante un ingreso, o predicciones de supervivencia de un paciente tras un diagnóstico de alta morbimortalidad (144).

En hematología se han diseñado aplicaciones de IA en las Neoplasias Mieloproliferativas Filadelfia negativas utilizadas para apoyar el diagnóstico, así como orientadas al pronóstico. Entre ellas encontramos algunas donde se predice el riesgo de progresión a Mielofibrosis, de residencia a tratamientos con la Hydroxiurea® o el riesgo de tener un evento tromboembólico (108).

En Mieloma múltiple se han publicado este mismo año 2023 los resultados de un algoritmo basado en aprendizaje automático supervisado que mejora la predicción del Sistema Internación de Supervivencia revisado (R-ISS), mejorando la predicción de supervivencia y respuesta al tratamiento en estos pacientes (145).

Por su lado, el grupo europeo de trasplante de medula ósea (EBMT) también ha utilizado la IA para predecir la mortalidad a los 100 días de trasplante, mediante técnicas de minería de datos en más de 26.000 pacientes, con resultados que mejoran los sistemas pronósticos actuales con curvas ROC que muestra un AUC de 0,701 (146).

Recientemente la sociedad americana de hematología (ASH) ha reconocido la validez de la Inteligencia Artificial para la clasificación morfológica de enfermedades hematológicas, proponiendo la participación de estos nuevos sistemas informáticos en las próximas actualizaciones de las clasificaciones de enfermedades (147).

La IA se ha utilizado también en el campo de los síndromes mielodisplásicos (SMD) para mejorar las asociaciones entre alteraciones cito-morfológicas y los estudios citogenéticos. La valoración cito-morfológica medular de esta patología es reconocida como una de las más complicadas y la experiencia del hematólogo juega un papel imprescindible para conseguir descripciones celulares y diagnósticos correctos. El objetivo final es poder prever alteraciones genéticas con mayor fiabilidad tras una primera valoración morfológica, ya que las dianas terapéuticas cada vez adquieren mayor importancia en el tratamiento de estas enfermedades (148). A este respecto, también se ha utilizado la IA para definir aquel paciente con mayor probabilidad de portar alteraciones genéticas de interés y, por tanto, candidato a estudios de secuenciación masiva, con el fin de disminuir los costes que suponen estos estudios y dirigirlos de mejor manera(149).

Los servicios transfusionales también se han apoyado en la IA en su búsqueda de mejoras en la política y actividades transfusionales, donde su participación ha demostrado gran utilidad en los procedimientos automáticos, así como en las estimaciones de consumo de componentes sanguíneos en cada centro hospitalario, lo que mejora la utilización de los recursos que disponemos, minimizando el desperdicio de los componentes hemáticos (150).

Podemos encontrar artículos que tratan de crear un fenotipo de paciente con tendencia a la depresión en aquellos pacientes con talasemia transfusión dependiente apoyados por sistemas de aprendizaje automático, así como artículos donde se proponen encontrar los factores de riesgo sociodemográficos más relevantes en mujeres o recién

nacidos para predecir la anemia llegando a obtener precisiones de hasta el 82% mediante Random Forest y técnicas de Boosting (151–154).

La IA está presente también en aquellos artículos donde no es el objetivo principal demostrar su validez. Trabajos con objetivos tan interesantes como la búsqueda de nuevas aplicaciones terapéuticas para fármacos ya utilizados en hematología, mediante la comparación de las alteraciones genéticas similares en diferentes enfermedades, están apoyadas en IA (155).

Para finalizar este punto, la IA también puede ser utilizada para conseguir prever eficacia en vida real de fármacos que han demostrado su eficacia en los ensayos clínicos, donde sabemos que en la mayoría de los casos tienen una población muy delimitada que no representa la población de pacientes que vemos en la práctica clínica, así como para valorar posibles respuestas a diferentes combinaciones terapéuticas de fármacos ya existentes, basándose en sus mecanismos de acción y posibles efectos coadyuvantes (156,157).

### 3.4 Inteligencia Artificial en la clasificación de anemias

Para tener una perspectiva real del interés de la IA por las anemias, debemos regresar hasta la década de los 80 y ver que ya se comenzaban a utilizar sistemas algoritmos de Machine Learning tradicionales con intención de guiar al médico (internista) en la batería de pruebas que debía escoger, según los resultados que iban obteniendo de las pruebas ya realizadas (iniciándose el estudio con un hemograma). Además de datos analíticos, en segundas etapas se iban incluyendo datos de la historia clínica y exploración física. Aunque la muestra de pacientes era limitada (320 pacientes) y no existía validación externa del sistema, los resultados publicados mostraban concordancia en el 90% de los resultados, cuando se comparan frente a un hematólogo especialista (8,158).

En el año 2023 encontramos artículos que continúan trabajando en la misma línea de investigación, con intención de mejorar la capacidad de clasificación de una maquina frente a los diferentes tipos de anemia, incluso en algunos casos llegamos a disponer de accesos webs a los algoritmos de clasificación para que se pueda acceder a él desde todas las zonas rurales o con menos capacidad económica(159).

### 3.4.1. Diagnóstico de anemias mediante ML tradicional en el análisis de datos tabulados

Desde hace varias décadas son bien conocidas las fórmulas matemáticas para el diagnóstico y/o screening de talasemias, sobre todo si hablamos de beta-talasemias y estas fórmulas han servido como objetivo a batir por los sistemas de IA. Existen publicados sistemas basados en SVM con los que se consiguen Valores Predictivos Negativos superiores al 99%, además de demostrar superioridad en resultado frente a más 25 fórmulas matemáticas publicadas en la bibliografía científica. Estos resultados tan llamativos son debidos a la sencillez que supone un problema de clasificación binaria (talasemias vs sanos). Además, estos resultados están respaldados en una gran cantidad de datos recopilados, que como ya hemos comentado mejora la capacidad de aprendizaje del sistema, utilizado en esta tesis unas 22.000 muestras (de las cuales el 13% eran portadoras de una beta-talasemia), además del trabajo de diseño y clasificación que existe detrás de los datos (160).

### 3.4.2. Diagnóstico de anemias mediante redes neuronales en el análisis de datos tabulados

Podemos encontrar artículos al respecto desde 1996, con un nivel de capas neuronales reducido, pero con un buen número de datos para el entrenamiento del sistema, en los que la máquina es capaz de discernir entre 3 tipos de anemias (ferropénica, hemoglobinopatías y trastorno crónico) con una precisión del 96%. En este sentido, los mismos autores comentan que la ausencia de validación externa hace que sus resultados pierdan fiabilidad (161).

Como hemos visto en secciones anteriores, existen algoritmos con múltiples capas neuronales, basados en hemogramas y estudios genéticos, así como raza y sexo, que han demostrado mejorar los sistemas de screening utilizados en talasemias hoy en día.

El artículo publicado en IA-hematología con mayor número de muestras utilizadas (8.693) para el aprendizaje profundo de una máquina se ha utilizado para el diagnóstico de hemoglobinopatías, en concreto para el despistaje de talasemias. Las curvas ROC publicadas por el grupo de *Mo, Donghua et al.* muestran un área bajo la curva (AUC) de 0,96 y, por tanto, mejoran hasta un 10% al sistema de screening de la población donde se realiza el estudio: Guangzhou, China (162).

Tanto si hablamos de trabajos basados en el aprendizaje de redes neuronales, como si hablamos de métodos tradicionales de aprendizaje automático, el objetivo de la práctica totalidad de ellos es conseguir discernir entre muestras de pacientes con talasemia y ferropenia, así como para el uso nuevos sistemas de *Screening*. Es decir, la mayoría de ellos trabaja en conseguir una correcta clasificación binaria (163–172) , algo que ya se

había buscado conseguir con la utilización de las fórmulas matemática, el objetivo a batir (173,174).

### 3.4.3. Diagnóstico de anemia mediante el análisis de imágenes

En el capítulo 1.3.2 ya hemos comentado las posibilidades que ofrece la IA en el diagnóstico de imágenes hematológicas y en esta misma línea encontramos sistemas como el propuesto por *Shir Ying et al.* basados en redes neuronales convolucionales en las que a partir de la detección de inclusiones eritrocitarias de HbH, son capaces de detectar alfa talasemias, con una sensibilidad y especificidad del 91 y 99% respectivamente (175).

Encontramos también aplicaciones inteligentes dedicadas al análisis de imágenes de electroforesis con precisiones superiores al 95% en el diagnóstico de hemoglobinopatías (176).

Si nos referimos a aquellas anemias congénitas de menor frecuencia, también encontramos algunos algoritmos capaces de predecir la alteración genética de un paciente basándose en la estructura celular 3D del glóbulo rojo. Es el caso de los pacientes con esferocitosis hereditaria, donde, según el perfil de la membrana observado por la máquina, se intenta conocer la mutación concreta de la enfermedad (177).

#### 3.4.4. Otras formas de aplicación de Inteligencia Artificial en el estudio de anemias

Otro punto en el que se trabaja en la medicina personalizada y de precisión es en posibilidad de minimizar los procesos invasivos. Encontramos esta rama de la IA para el diagnóstico de anemias con aplicaciones que evitan la extracción sanguínea y es que existen estudios en los que el diagnóstico está basado en una pletismografía, observándose correlaciones de hasta el 81% entre los datos de la pletismografía y el hemograma control (178).

En esta misma línea de trabajo encontramos aplicaciones de diagnóstico de anemia basados en imágenes de la conjuntiva palpebral, utilizando imágenes de fondo de ojo para entrenar al sistema inteligente, validado con pacientes con diabetes mellitus, que precisan frecuentes revisiones oftalmológicas (179–181).

También vamos a encontrar sistemas entrenados con fotografías faciales, utilizando la cámara del *Smartphone*, consiguiendo sensibilidades del 90-94% (182,183) o diagnóstico de anemia basados en electrocardiogramas. En este último trabajo se consiguieron sensibilidades entre el 83-86%, en un trabajo que aporta validación interna y externa, lo cual aporta mayor calidad a sus resultados (184).

Al igual que los entrenamientos basados en imágenes, los estudios basados en el reconocimiento de lenguaje natural están en pleno apogeo. En este sentido encontramos publicados artículos en los que encontramos diagnósticos de anemias basados en los comentarios publicados en redes sociales con precisiones superiores al 80%, aunque sin validación externa (185).

En definitiva, el diagnóstico y screening de anemias, tanto congénitas como adquiridas está en un proceso de cambio y mejora constante. La IA y los nuevos sensores digitales están facilitando el acceso y mejorando los resultados de diagnósticos. Uno de los últimos artículos publicados, plantea la opción de clasificar aquellos pacientes con anemias congénitas que precisan realmente un manejo hospitalario vs fenotipo asintomático que probablemente no precise ni seguimiento analítico (186).

### 3.5 Limitaciones de la Inteligencia Artificial en salud

Probablemente el bombo de la Inteligencia Artificial, muy relacionada con la ciencia ficción, supera en la actualidad el estado del arte de la IA en medicina (76). Ejemplo de ello son algunas promesas como la del sistema “*Watson for Oncology*”, un algoritmo de IA creado por IBM® para el apoyo en el diagnóstico de tratamiento de enfermedades oncológicas utilizado por cientos de hospitales hasta 2019, cuando se publicaron importantes errores en sus recomendaciones de tratamientos, cerrándose el proyecto en la mayoría de sus centros. A día de hoy se continua utilizando en algunos países y, como en la mayoría de algoritmos de IA, la limitada base de datos de la que disponían en 2019 fue la gran limitante en sus recomendaciones (187,188). La solución que se propone desde algunos grupos de investigación es la de introducir en los hospitales bases de datos racionales, es decir, sistemas informáticos que vayan relacionando diagnósticos con cuidados y tratamientos en el día a día. *BRAIN Project* trata de hacer esto con datos recogidos en los servicios de hematología y oncología del Hospital Universitario de Heidelberg (Alemania) desde 2008. De esta manera se disminuye el error en transcripción. Estos sistemas, una vez más, están basados en reconocimiento del lenguaje natural asociado con *Big Data* (189).

Aunque no es una limitación de la IA, otro problema que plantea la entrada de estos sistemas en el sistema sanitario es que la automatización de los diagnóstico pueda disminuir la destreza del médico del futuro. Además de que la pérdida del contexto del diagnóstico o la incertidumbre en el diagnóstico médico (refiriéndose a las diferentes valoraciones u opiniones iniciales sobre un diagnósticos complicado), es una situación

frecuente en los hospitales, ya que en medicina no todo entra dentro de los patrones diagnósticos de los libros (190).

Por otro lado, existe una ausencia de ensayos clínicos con datos prospectivos para el aprendizaje del sistema inteligente, lo que puede dificultar su uso en el ámbito médico, aunque está en proceso de solucionarse (105). De hecho, ya hemos comentamos como las guías CONSORT (utilizadas para la correcta publicación de ensayos clínicos) se han adaptado y han publicado su actualización para ensayos en los que se incluyan algoritmos de Inteligencia Artificial: las guías CONSORT-IA (71).

El tema más polémico y que más está limitando a la IA es la capacidad de explicabilidad de los procesos ocultos utilizados hasta conseguir el resultado obtenido. Se conoce como La Caja Negra o *Black Box* a los procesos internos (sobre todo cuando se habla de *Deep Learning*) en los que los sistemas de clasificación utilizados no quedan explicados y, por tanto, disminuyen la confianza que aportan estos sistemas para una aplicabilidad en sistema sanitario (191). Es decir, vemos como la maquina recoge todos los datos y consigue obtener un resultado, pero no conocemos cómo lo hace y realmente es un tema que aún sigue abierto a debate. Por ello, la investigación en esta línea es objetivo imprescindible si se quiere mejorar la confianza en estos sistemas (192). Existen trabajos extensos que quedan lejos de nuestro conocimiento como sanitarios, pero muy enfocados en conseguir un objetivo común con el sanitario: darle explicabilidad a la clasificación inteligente automatizada. Probablemente la manera de evaluar esta explicabilidad es un tema que debemos dejar de la mano de ingenieros informáticos pero no debemos despreciarla cuando se nos ofrezca y entenderla en su espectro más básico (193).

La causabilidad y explicabilidad son objeto de publicaciones, en las que nos explican cómo los diferentes pesos de las variables van cambiando para que los resultados de clasificación vayan mejorando. Como posible solución a este problema, publicaciones recientes nos muestran cómo se pueden extraer datos para conocer cuáles son las variables en las que la maquina se ha apoyado en mayor o menor medida para conseguir la clasificación de los datos en cuestión (194,195).

Para finalizar esta introducción, debemos recordar que una de las cosas más importantes para un paciente y un médico es la relación creada entre ambos. El paciente aprecia en gran medida la posibilidad de poder expresar ante su médico sus dolencias. Incluso lo hemos vivido durante la pandemia COVID-19 y las llamadas telefónicas, siendo difíciles de gestionar en muchos casos por la necesidad de valoración presencial por algunos pacientes. Debemos informatizarnos y modernizarnos, pero sin olvidar que el paciente necesita contacto humano(196).

## 4 Hipótesis y objetivos

La anemia es una de las afecciones más frecuentes y, como hemos visto, su prevalencia global llega al 24,8%, afectando a más de 1,5 billones de personas en el mundo. Asimismo, debido al alto flujo migratorio, las hemoglobinopatías se posicionan entre sus causas más frecuentes en todos los territorios.

Por otro lado, la Inteligencia Artificial, una rama de las ciencias de la computación, ha llegado a nuestras vidas en forma de diversas aplicaciones digitales, la mayoría de ellas proyectadas sobre sistemas inteligentes como relojes, móviles, televisores o futuros coches de conducción autónoma. Asimismo, hemos visto como en los últimos 10 años la participación de la IA en medicina (y en hematología en concreto) es una realidad, marcándose objetivos como la mejora en los tiempos de espera, la precisión y velocidad de diagnóstico, la disminución de errores y la mejora en la gestión de recursos. Los resultados publicados hasta la fecha son muy prometedores, sin embargo, el desconocimiento en cuanto a la explicabilidad de estos sistemas inteligentes sigue siendo una caja negra difícilmente decodificable que mantiene escépticos a una parte del mundo sanitario. Es por ello que una gran parte de los actuales trabajos van dirigidos a aclarar estas dudas.

La hipótesis de la presente tesis pretende demostrar que es posible entrenar un sistema inteligente para que, a partir de hemogramas con diagnósticos de anemias de diferente origen (incluyendo hemogramas de pacientes sanos como control) y mediante modelos complejos de aprendizaje profundo (Redes Neuronales), sea capaz de realizar un diagnóstico de precisión, como se ha demostrado en otros campos de la hematología.

Para validar esta hipótesis nos hemos planteado los siguientes objetivos:

- 1) Investigar y desarrollar un sistema de diagnóstico diferencial de anemias basado en Inteligencia Artificial (Redes Neuronales) a partir de muestras de hemogramas. Para ello nos hemos planteado varios objetivos secundarios:
  - Aumentar la complejidad del problema de clasificación de anemias con respecto a lo publicado en el estado del arte sin perder precisión en el resultado.
  - Demostrar que la precisión y exactitud del sistema se verán incrementadas de manera paralela al número de muestras de hemogramas y que, por el contrario, esta precisión disminuirá con el aumento del número de diagnósticos, al aumentar la complejidad del problema.
  - Certificar la importancia de la calidad de los datos que se utilizaran en el entrenamiento de las Redes Neuronales.
  
- 2) Interpretar y entender el método de clasificación llevado a cabo por nuestra Red Neuronal hasta obtener los resultados. Para ello nos hemos planteado los siguientes objetivos secundarios:
  - Realizar un análisis de los resultados ofrecidos por nuestra Red Neuronal mediante programas de explicabilidad ya validados previamente por la comunidad científica (valores SHAP).
  - Identificar cuáles son las variables de mayor peso y los valores de las mismas que permiten diferenciar entre las distintas anemias.

3) Confirmar la validez y la utilidad del sistema de diagnóstico de Inteligencia Artificial desarrollado. Para ello nos planteamos varios objetivos secundarios:

- Realizar estudios comparativos con otros métodos de clasificación tradicionales de aprendizaje de máquinas (Machine Learning) en la búsqueda de mejorar los niveles de precisión y exactitud reportados en la literatura científica actual.
- Confirmar los resultados del sistema en otro grupo de pacientes que no haya tenido contacto con el sistema de aprendizaje automático durante su entrenamiento.
- Evaluar los posibles beneficios de coste-eficacia en la aplicación de nuestra Red Neuronal.

# 5. Materiales y Métodos

## 5.1 Búsqueda bibliográfica

### 5.1.1 Estrategia de búsqueda

Se ha realizado una búsqueda bibliográfica en la literatura científica en las bases de datos internacionales PUBMED-MEDLINE, Web Of Science (WOS), EMBASE y Virtual Health Library (VHL = BVS). También se ha realizado la búsqueda bibliográfica en repositorios de origen nacional como Dialnet. De la misma manera hemos realizado una búsqueda de tesis doctorales nacionales/internacionales en TESEO, TDR, DART-Europe, Proquest, OhioLink, Cybertesis, TEL y Networked Digital Library. Además, hemos añadido información de páginas web, hemos utilizado Google Scholar y hemos realizado búsqueda manual. Se han revisado artículos escritos en inglés y español (castellano y latino) sin limitación en los años de búsqueda.

Hemos utilizado el programa Mendeley Desktop® para la gestión, organización e integración de todas las fuentes bibliográficas. Todas estas referencias conforman el apartado *bibliografía* y han sido citadas según el estilo *Vancouver*.

Hemos localizado un total de 4.696 publicaciones asociadas de alguna manera al diagnóstico de anemias mediante la utilización de IA, repartidos de la manera que aparece en la **Tabla 1**. Tras una primera selección de artículos relacionados y una segunda de artículos relevantes, finalmente y tras una tercera selección hemos seleccionado un total de 214 artículos, revisiones, comunicaciones a congresos o dominios webs, que se han utilizado para describir el estado de arte de este tema.

Inicialmente definimos los términos claves en lenguaje libre del tema a abordar: Hemogramas, anemia, talasemia, hemoglobinopatías, herramienta diagnóstica, diagnóstico, Inteligencia Artificial, aprendizaje automático, redes neuronales, redes neuronales artificiales, máquinas de soporte vectorial, bosques aleatorios y árboles de decisión.

<b>BASE DE DATOS</b>	<b>Artículos recuperados</b>	<b>Artículos relacionados</b>	<b>Artículos relevantes</b>	<b>Artículos referenciados</b>
PUBMED	1281	195	98	69
WOS	1319	165	59	35
EMBASE	1782	107	67	32
Dialnet	169	16	9	3
<b>Portal Regional de la BVS:</b>	67	15	2	2
IBECS	14	1	0	0
BDENF	3	1	0	0
LILACS	37	11	2	2
CUMED	6	1	0	0
INDEXPSI	2	0	0	0
MULTIMEDIA	2	0	0	0
SES-SP	2	0	0	0
BINACIS	1	1	0	0
<b>Tesis Doctorales:</b>	6	4	3	0
TESEO	0	0	0	0
TDR	1	1	1	0
DART-Europe	2	1	0	0
OhioLINK	2	1	1	0
TEL	0	0	0	0
PROQUEST	0	0	0	0
CYBERTESIS	1	1	1	0
Networked Digital Library (NDLTD)	0	0	0	0
Búsqueda manual	76	76	76	74
<b>TOTAL</b>	4696	574	310	215

TABLA 1. RESULTADOS, CLASIFICACIÓN Y SELECCIÓN DE LA BÚSQUEDA BIBLIOGRAFÍA REALIZADA.

Algunos de estos conceptos se transformaron a términos del lenguaje controlado y se procedió a construir la estrategia de búsqueda con los correspondientes booleanos de cada buscador, como se muestra a continuación, según cada buscador utilizado.

### 5.1.2 Base de datos PUBMED (National Library of Medicine).

Se ha utilizado la siguiente estrategia de búsqueda bibliográfica:

("random decision forest"[tiab] OR "Decision Trees"[Mesh] OR "Random Forest"[Mesh] OR "Support Vector Machine"[Mesh] OR "Learning Machine"[tiab] OR "Transfer Learning"[tiab] OR "Learning Transfer"[tiab] OR "Machine Learning"[Mesh] OR "Intelligence Artificial"[tiab] OR "Computational Intelligence"[tiab] OR "Intelligence Computational"[tiab] OR "Machine Intelligence"[tiab] OR "Intelligence Machine"[tiab] OR "Computer Reasoning"[tiab] OR "Reasoning Computer"[tiab] OR "AI (Artificial Intelligence)"[tiab] OR "Computer Vision System\*"[tiab] OR "System Computer Vision"[tiab] OR "Systems Computer Vision"[tiab] OR "Vision System Computer"[tiab] OR "Vision Systems Computer"[tiab] OR "Knowledge Acquisition (Computer)"[tiab] OR "Acquisition Knowledge (Computer)"[tiab] OR "Knowledge Representation (Computer)"[tiab] OR "Knowledge Representations (Computer)"[tiab] OR "Representation Knowledge (Computer)"[tiab] OR "Artificial Intelligence"[Mesh] OR "Artificial intelligence"[tiab] OR "machine learning"[tiab] OR "Artificial neuronal networks"[tiab] OR "Neuronal networks"[tiab] OR "Support vector machine"[tiab] OR "Random forest"[tiab] OR "Decision trees"[tiab] OR "Neural Networks, Computer"[Mesh]) **AND** (Anemia[tiab] OR Anaemia[tiab] OR Thalassemia[tiab] OR Hemoglobinopathy[tiab] OR "hematologic diseases"[tiab] OR "Hematologic Diseases"[Mesh] OR "hematological diseases"[tiab] OR "haematological diseases"[tiab] OR "blood diseases"[tiab] OR hematology[tiab] OR haematology[tiab])

### 5.1.3 Base de datos EMBASE

Se ha utilizado la siguiente estrategia de búsqueda bibliográfica:

('random decision forest':ti,ab OR 'decision tree'/exp OR 'random forest'/exp OR 'support vector machine'/exp OR 'learning machine':ti,ab OR 'transfer learning':ab,ti OR 'learning transfer':ti,ab OR 'machine learning'/exp OR 'intelligence artificial':ab,ti OR 'computational intelligence':ab,ti OR 'intelligence computational':ab,ti OR 'machine intelligence':ab,ti OR 'intelligence machine':ab,ti OR 'computer reasoning':ab,ti OR 'reasoning computer':ab,ti OR 'ai (artificial intelligence)':ab,ti OR 'computer vision system\*':ab,ti OR 'system computer vision':ab,ti OR 'systems computer vision':ab,ti OR 'vision system computer':ab,ti OR 'vision systems computer':ab,ti OR 'knowledge acquisition (computer)':ab,ti OR 'acquisition knowledge (computer)':ab,ti OR 'knowledge representation (computer)':ab,ti OR 'knowledge representations (computer)':ab,ti OR 'artificial intelligence'/exp OR 'artificial intelligence':ab,ti OR 'machine learning':ab,ti OR 'artificial neuronal networks':ab,ti OR 'neuronal networks':ab,ti OR 'support vector machine':ab,ti OR 'random forest':ab,ti OR 'decision trees':ab,ti OR 'artificial neural network'/exp) **AND** ('anemia'/exp OR 'anemia':ab,ti OR 'anaemia':ab,ti OR 'thalassemia':ab,ti OR 'hemoglobinopathy':ab,ti OR 'hematologic disease':ab,ti OR 'hematological diseases':ab,ti OR 'haematological diseases':ab,ti OR 'blood diseases':ab,ti OR 'hematology':ab,ti OR 'haematology':ab,ti OR 'hematologic disease'/de) **AND** [embase]/lim

#### 5.1.4 Base de datos WOS (Web Of Sciences)

Se ha utilizado la siguiente estrategia de búsqueda bibliográfica:

(Artificial intelligence OR Neuronal Networks OR decision trees) (Topic) **AND**  
(hematology OR anemia OR thalassemia OR hemoglobinopathy) (Topic)

#### 5.1.5 Base de datos Dialnet

Inteligencia Artificial **Y** (análisis clínicos O sangre O anemia O diagnóstico O hematología  
O hemograma).

#### 5.1.6 Portal Regional de la BVS

Se ha utilizado la siguiente estrategia de búsqueda bibliográfica:

Inteligencia Artificial OR redes neuronales OR arboles de decisión) **AND** (anemia OR  
hematología OR talasemia OR hemoglobinopatía OR diagnóstico)

#### 5.1.7 Tesis Doctorales

Se ha utilizado la siguiente estrategia de búsqueda bibliográfica:

(Artificial intelligence OR Neuronal Networks OR decision trees) **AND** (hematology OR  
anemia OR thalassemia OR hemoglobinopathy OR diagnosis).

## 5.2 Tipo de estudio y fuente de datos

Estudio de cohortes retrospectivo unicéntrico. Para la primera parte del trabajo (entrenamiento del sistema), se han recogido todos los hemogramas registrados en la base de datos del HCSC de Madrid entre los años 2011 y 2018, obteniéndose un total de 111.144 hemogramas de los cuales etiquetamos manualmente 4.060.

De los 4.060 hemogramas etiquetados, se han utilizado el 90% para el entrenamiento del sistema y un 10% para la validación. Para la validación externa o puesta a punto del sistema se han recogido todos los hemogramas registrados desde enero del 2019 a marzo del 2021 y se han etiquetado un total de 250 muestras.

A diferencia de lo que ocurre con la estadística clásica, para el diseño de Redes Neuronales no existen bases en cuanto a la estimación de la muestra. Por ello, la estimación de la muestra necesaria (hemogramas etiquetados) para alcanzar resultados óptimos en esta tesis está basada en aquellos trabajos realizados en este campo de la hematología con resultados favorables que se encontraron durante la revisión bibliográfica (117,129,197,198).

La población de estudio se conformó por todos los pacientes que se hubieran realizado un hemograma en el servicio de hematología del HCSC entre 2011 y 2018. La distribución de los pacientes a los que el HCSC le corresponde como centro de referencia, divididos por edades y los diferentes centros de salud dependientes del HCSC se muestra en la **Tabla 2**, según aparece publicado por el Ministerio de Sanidad de la Comunidad de Madrid (199).

NOMBRE CENTRO	LOCALIDAD	POBLACIÓN (01/01/18)					TOTAL
		0-2	3-15	16-64	65-79	>=80	
C.S. CEA BERMÚDEZ	MADRID	472	2.112	15.811	3.211	1.892	23.498
C.S. GUZMÁN EL BUENO	MADRID	450	1.786	14.652	2.996	1.748	21.632
C.S. ESPRONCEDA	MADRID	969	4.043	26.275	5.117	2.859	39.263
C.S. ELOY GONZALO	MADRID	716	3.317	23.482	4.304	2.659	34.478
C.S. CARAMUEL	MADRID	521	2.407	16.100	2.580	1.793	23.401
C.S. PUERTA DEL ÁNGEL	MADRID	364	2.104	13.615	2.250	1.771	20.104
C.S. LUCERO	MADRID	808	4.632	25.814	4.845	3.366	39.465
C.S. MAQUEDA	MADRID	530	3.045	17.431	5.220	2.349	28.575
C.S. LOS CÁRMENES	MADRID	501	2.864	13.587	1.870	1.569	20.391
C.S. CAMPAMENTO	MADRID	326	1.801	9.756	1.780	1.116	14.779
C.S. LAS ÁGUILAS	MADRID	592	2.868	16.884	4.367	2.313	27.024
C.S. VALLE INCLÁN	MADRID	219	1.188	6.113	1.755	983	10.258
C.S. GENERAL FANJUL	MADRID	488	2.636	14.808	3.630	1.242	22.804
C.S. GENERAL RICARDOS	MADRID	1.053	5.521	32.753	5.650	3.632	48.609
	<b>TOTAL</b>	<b>8.009</b>	<b>40.324</b>	<b>247.081</b>	<b>49.575</b>	<b>29.292</b>	<b>374.281</b>

TABLA 2. POBLACIÓN ESTUDIADA, DIVIDIDA POR CENTROS DE SALUD DEPENDIENTES DEL HOSPITAL CLÍNICO SAN CARLOS. LAS COLUMNAS ESTÁN DIVIDIDAS POR FRANJAS DE EDAD.

Tras la aplicación de los criterios de exclusión y la limpieza de datos, llevada a cabo tal y como se explica en el apartado 5.6, entrenamos al sistema con un total de 3.295 hemogramas que se distribuyen según se refleja en la **Tabla 3**.

Etiqueta	Numero de muestras
Sanos	863
Talasemias	675
Hemoglobinopatías estructurales	798
Ferropénicas	455
Trastorno Crónico	504
<b>TOTAL</b>	<b>3295</b>

TABLA 3. MUESTRAS INCLUIDAS EN EL APRENDIZAJE DE BLOODNET, DISTRIBUIDAS POR SUS DIAGNÓSTICO.

### 5.3 Variables incluidas en el entrenamiento del algoritmo

Se han incluido 17 variables de cada muestra procesada: 15 variables analíticas del hemograma, representadas en la **Tabla 4** así como dos variables no analíticas (edad y el sexo).

Abreviatura	Variable (unidad de medida)	Valores referencia			
		Femenino		Masculino	
		LI	LS	LI	LS
HEM	Hematíes ( $10^6/\mu\text{L}$ )	3,9	5,2	4,2	5,6
HB	Hemoglobina (g/dL)	119	154	131	172
HTO	Hematocrito (%)	35	46	39	59
VCM	Volumen Corpuscular Medio fL	81	-	-	101
MCH	Hemoglobina Corpuscular Media (pg)	27	-	-	35
CHCM	Concentración corpuscular media de Hemoglobina (g/dL)	31	35	32	36
ADE	Amplitud de distribución de los hematíes (%)	12,5	-	-	14,5
WBC	Leucocitos totales ( $10^3/\mu\text{L}$ )	4	-	-	10
LIAB	Valor absoluto de Linfocitos( $10^3/\mu\text{L}$ )	1	-	-	3
MOAB	Valor absoluto de Monocitos( $10^3/\mu\text{L}$ )	0,2	-	-	1
NEAB	Valor absoluto de Neutrófilos( $10^3/\mu\text{L}$ )	2	-	-	7
EOAB	Valor absoluto de Eosinófilos ( $10^3/\mu\text{L}$ )	0,02	-	-	0,5
BAAB	Valor absoluto de Basófilos ( $10^3/\mu\text{L}$ )	0,02	-	-	0,1
PLT	Plaquetas ( $10^3/\mu\text{L}$ )	150	-	-	450
VPM	Volumen plaquetario medio (fL)	7,4	-	-	11,4

**TABLA 4. VARIABLES Y VALORES REFERENCIA UTILIZADOS POR EL ANALIZADOR *BECKMAN COULTER-DXH*. LI: LÍMITE INFERIOR. LS LÍMITE SUPERIOR.**

## 5.4 Criterios de inclusión

Se han incluido para el entrenamiento los hemogramas de pacientes con diagnóstico de hemoglobinopatía (talasemias y hemoglobinopatías estructurales), anemia ferropénica, anemia por trastorno crónico/inflamatorio y hemogramas de pacientes sanos. Todos los subgrupos de hemoglobinopatías incluidos en el estudio, además del resto de anemias, están representados en la **Tabla 5**.

<b>1</b>	<b>Anemias adquiridas</b>
1.1	ANEMIA FERROPENICA
1.2	ANEMIA POR TRANSTORNO CRÓNICO
<b>2</b>	<b>Anemias congénitas</b>
	Hemoglobinopatías estructurales
2.1	HEMOGLOBINA C HETEROCIGOTA
2.2	HEMOGLOBINA C HOMOCIGOTA
2.3	HEMOGLOBINA S HETEROCIGOTA
2.4	HEMOGLOBINA S HOMOCIGOTA
2.5	HEMOGLOBINA E
2.6	HEMOGLOBINA D (LOS ANGELES)
2.7	OTRA HEMOGLOBINOPATIAS
	Talasemias
2.8	ALFA TALASEMIA (-+/++)
2.9	ALFA TALASEMIA (--/++)
2.10	ALFA TALASEMIA (+-/+-)
2.11	ALFA TALSEMIA (--/-+) (ENFERMEDAD HEMOGLOBINA H)
2.12	BETA 0 TALASEMA Heterocigota
2.13	BETA + TALASEMIA heterocigota
2.14	BETA + TALASEMIA homocigota
2.15	DELTA-BETA 0 TALASEMIA HETEROCIGOTA (SPANISH TALASEMIA)
2.16	EPSILON-GAMMA-DELTA-BETA TALASEMIA
<b>3</b>	<b>Hemogramas control (pacientes sanos)</b>

TABLA 5. DIAGNÓSTICOS DIVIDIDOS POR SUBGRUPOS DE LAS MUESTRAS DE HEMOGLOBINOPATÍAS INCLUIDAS EN EL ENTRENAMIENTO DEL ALGORITMO INTELIGENTE

## 5.5 Sistemas de diagnóstico hematológico

A continuación, se describen los métodos de trabajo y diagnóstico seguido en el Hospital Clínico San Carlos (HSCS) para alcanzar el diagnóstico final de los pacientes con anemia y hemoglobinopatías, centrado en este último punto que alberga mayor complejidad.

Todos los hemogramas se han procesado mediante el analizador Beckman Coulter-DxH®. Las anemias ferropénicas y por trastorno crónico se diagnosticaron en base al hemograma, el frotis de sangre periférica y la bioquímica, donde se incluía hierro, ferritina, transferrina e índice de saturación de la transferrina.

En cuanto a las hemoglobinopatías, se utilizó la Cromatografía Líquida de Alta Resolución (HPLC, de sus siglas en inglés) para medir los niveles de HbA<sub>2</sub> y Hb Fetal. Las Hemoglobinas fueron analizadas por electroforesis capilar zonal y HPLC de intercambio iónico. El ADN genómico fue procesado mediante un NanoDrop 1000 (Thermo Scientific, Wilmington, DE, EE. UU).

Para el diagnóstico de las  $\alpha$ -Talasemias se utilizó la PCR multiplex seguida de hibridación inversa para el cribado inicial y posteriormente los métodos de *Multiplex Ligation-dependent Probe Amplification* (MLPA) para su confirmación. El diagnóstico de las Talasemias no alfa se realizó mediante secuenciación automática mediante el método de *Sanger* (200).

Realizamos una recogida manual de datos de todas las muestras que se habían recibido en el HCSC durante el año el 2019 para estudio de Hemoglobinopatías. El objetivo de esta recogida de datos fue reflejar el elevado número de muestras recibidas en este centro anualmente y conocer el porcentaje de estas muestras que había acabado en diagnóstico y cuál había sido cada uno de los diagnósticos. Se recibieron un total de

1.113 muestras y fueron diagnosticadas 752 de hemoglobinopatías (67,5%), correspondiendo a los diagnósticos que se muestran en la **Tabla 6**.

DIAGNOSTICOS	MUESTRAS	% SOBRE EL TOTAL DE DIAGNOSTICOS
<b>Talasemias</b>	<b>434</b>	<b>57,71</b>
Alfa talasemias	284	37,76
Alfa + heterocigotas	183	24,3
alfa + homocigotas	64	8,51
Alfa0 heterocigotas	32	4,25%
Hb H	5	0,66
Beta talasemias	126	16,75
Delta-Beta talasemia ( <i>Spanish</i> talasemia)	24	3,19
<b>Hemoglobinopatías estructurales</b>	<b>318</b>	<b>42,28</b>
Hemoglobinopatía C	40	5,31
Hemoglobinopatía S	179	23,80
Hemoglobinopatía E	7	0,93
Hemoglobinopatía D	8	1,06
Otras Hemoglobinopatías	84	11,17
<b>TOTAL</b>	<b>752</b>	<b>100</b>

TABLA 6. DIAGNÓSTICOS DE HEMOGLOBINOPATÍAS LLEVADOS A CABO EN EL HCSC DE MADRID EN EL AÑO 2019

## 5.6 Criterios de exclusión: limpieza de datos.

Se excluyeron las siguientes muestras para el entrenamiento del algoritmo, con el fin de mejorar la calidad de datos (limpieza de datos):

- a) Los artefactos del hemograma producidos por conservación de la muestra, es decir, todas las muestras artefactadas al microscopio óptico, según criterio del médico hematólogo del HCSC.
- b) Todas las muestras de pacientes que hubiesen recibido transfusiones durante los 3 meses previos al estudio.
- c) Trombopenias con cifras inferiores a  $100 \times 10^3/\mu\text{L}$  plaquetas.
- d) Neutropenias inferiores a  $1.5 \times 10^3/\mu\text{L}$  neutrófilos.
- e) Cualquier causa de anemia que no se encuentre dentro de los criterios de inclusión.
- f) Presencia de 2 o más diagnósticos de manera simultánea en la misma muestra (p.e talasemia + ferropenia).
- g) Hemogramas con valores perdidos en las variables correspondientes a la serie roja.

Además, de manera digital se han filtrado aquellos valores estadísticamente atípicos. Para ello, se han eliminado los valores cuya diferencia de la media fuera superior a 6 veces la desviación típica. Estos valores se aplicaron por cada uno de los grupos diagnósticos, ya que los resultados de un paciente sano difieren de los de un paciente con talasemia por definición.

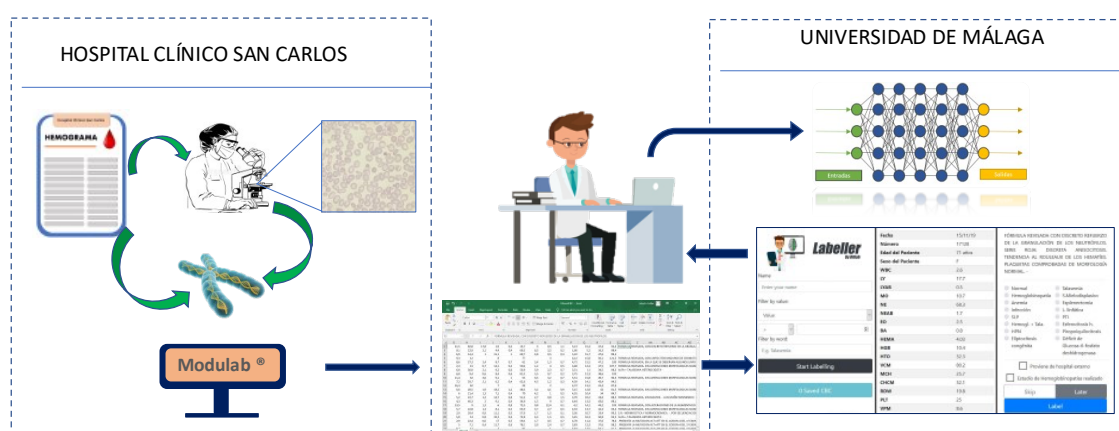
En cuanto a los valores perdidos o *missing values* son mínimos en esta tesis doctoral ya que al haber sido realizado el etiquetado manualmente por un hematólogo, no se etiquetaron los hemogramas con valores perdidos en la serie roja (criterio de exclusión).

Los escasos valores perdidos de hemogramas etiquetados corresponden a valores de la formula leucocitaria y se han completado con el valor medio de la población para dicha variable.

## 5.7. Digitalización de la muestra

Inicialmente se exportaron todos los hemogramas registrados entre 2011 y 2018 desde la base de datos del programa Modulab® del HCSC, obteniendo un documento Excel.

Se exportó el hemograma completo, los comentarios de los frotis de sangre periférica (texto libre redactado por el hematólogo correspondiente), el estudio genético si correspondía y los diagnósticos de cada muestra (**Figura 16**).



**FIGURA 16.** FLUJO DE TRABAJO ENTRE EL DEPARTAMENTO DE HEMATOLOGÍA DEL HCSC Y EL DEPARTAMENTO DE ARQUITECTURA DE LA COMPUTACIÓN DE LA UMA

A partir del documento Excel con los datos descritos previamente, se digitalizaron los hemogramas por parte del equipo informática de la UMA, mediante una herramienta web diseñada únicamente para este proyecto con utilidad en móviles y ordenadores que denominamos "Labeller". Esta aplicación conectaba los valores del hemograma, el estudio genético y el comentario del médico hematólogo sobre la extensión de sangre periférica, con una interfaz amigable para el usuario (**Figura 17**).

**Labeller**  
By Bitlab

Name: Enter your name

Filter by value: Value >

Filter by word: E.g. Talasemia

Start Labelling

0 Saved CBC

Fecha	25/09/19
Número	14125
Edad del Paciente	35 años
Sexo del Paciente	M
WBC	5
LY	38.5
LYAB	1.9
MO	6
NE	49.4
NEAB	2.5
EO	5.2
BA	0.9
HEMA	7.25
HGB	15.6
HTO	48.9
VCM	67.4
MCH	21.5
CHCM	31.9
RDW	15.8
PLT	171
VPM	8.2

FÓRMULA REVISADA SIN ALTERACIONES MORFOLÓGICAS SIGNIFICATIVAS. SERIE ROJA: MICROCÍTICA HIPOCROMA, ALGUNOS DIANOCITOS. PLAQUETAS COMPROBADAS DE MORFOLOGÍA NORMAL. - ALFA 0 TALASEMIA (-SEA/ALFA ALFA) -

- ALFA TALASEMIA (-+/+)
- ALFA TALASEMIA (+-/+-)
- ALFA TALASEMIA (-/-+)
- ENFERMEDAD HEMOGLOBINA H (-/-+)
- BETA 0 HETEROCIGOTA
- BETA + HETEROCIGOTA
- BETA + HOMOCIGOTA
- OTRAS TALASEMIAS
- SPANISH 0
- EPSILON GAMMA BETA DELTA T. HETEROCIGOTA
- ALFA TALASEMIA (-/-+)

Proviene de hospital externo

Estudio de Hemoglobinopatías realizado

Skip Later Label

FIGURA 17. INTERFACE DEL MÉTODO DE ETIQUETADO (LABELLER) DURANTE EL PROYECTO DE TESIS DOCTORAL

## 5.8. Procesamiento de datos: Enseñando a la máquina

A partir de los datos mostrados por la interface web digital *Labeller*, el hematólogo ha sido el encargado de etiquetar mediante el diagnóstico correspondiente cada uno de los hemogramas aplicando los criterios de inclusión comentados anteriormente. Una vez finalizado este proceso, se aplicaron los criterios de exclusión para confirmar con un experimento el valor de la calidad de datos en el aprendizaje automático inteligente (escenario 3.5).

Para poder localizar entre los más de 100.000 hemogramas recogidos aquellos correspondientes a los criterios de inclusión nos hemos servido de un buscador de palabras claves diseñado por el equipo de ingeniería informática (**Figura 17**). Este buscador rastrea las palabras claves elegidas a 3 niveles: en el texto libre del frotis de sangre periférica, en el texto del estudio genético y en el texto del diagnóstico final.

Las palabras claves utilizadas fueron: Anemia, trastorno crónico, ferropenia, ferropénica, microcítica, microcitosis, talasemia, hemoglobinopatía estructural, normocítica y normocrómica.

## 5.9. Métodos de Inteligencia Artificial aplicados

En esta tesis hemos aplicado métodos de aprendizaje automático supervisados, aportándole al sistema inteligente las etiquetas (diagnósticos) correspondientes a cada uno de los hemogramas. Inicialmente se han aplicado modelos tradicionales como las Máquinas de Soporte Vectorial, Random Forest y XGBoost, para posteriormente continuar con la aplicación de Redes Neuronales.

El equipo de Inteligencia Artificial del departamento de arquitectura de computadores de la Universidad de Málaga ha creado una Red Neuronal propia que hemos denominado BloodNet® y que hemos comparado con otra Red Neuronal de acceso público previamente diseñada por otro grupo de investigación (TabNet®).

BloodNet está basada en los mismos pilares del perceptrón multicapa (estructura básica de la Red Neurona similar a la reflejada en la **Figura 15**), adaptada a las características de los datos de esta tesis (hemogramas), siendo sus dos principales características:

- 1) Los datos de entrada se han separados por cada línea celular, es decir, inicialmente partimos de 3 capas principales: una capa de entrada para la serie eritrocitaria, otra capa de entrada para la serie leucocitaria y una tercera para la serie megacariocítica. Las variables de estas tres capas inicialmente interactúan entre sí y posteriormente lo hacen con el resto de variables. Este método aplicado está inspirado en la red COxPASNeT® en la cual interactúan datos clínicos y analíticos, pero que inicialmente lo hacen de manera separada (201). Esto es lo que conocemos como la incorporación del conocimiento clínico específico a nuestra Red Neuronal. Es decir, la inclusión de técnicas para suministrar conocimientos de lógica humana de un especialista hematólogo para

aprender a leer un hemograma, lo que ha sido nuestra mayor aportación científica.

2) Se ha proporcionado conocimiento a la red sobre los rangos de los valores normales de dichos grupos.

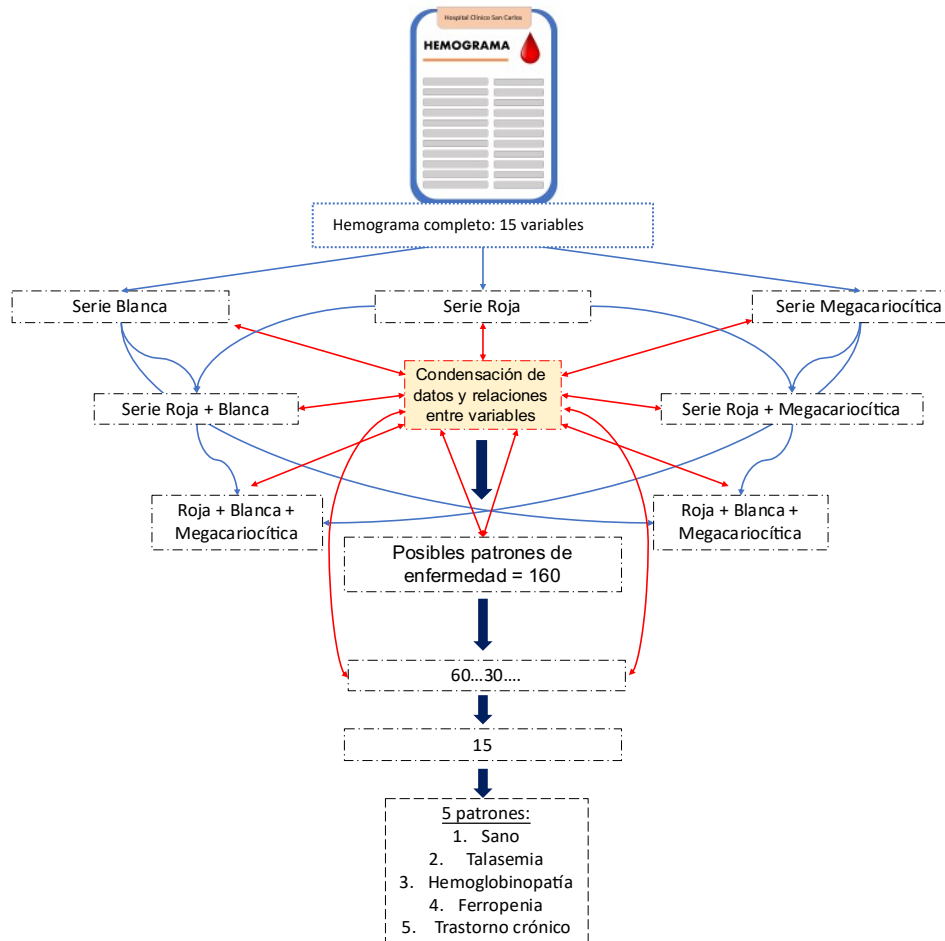


FIGURA 18. REPRESENTACIÓN DEL PROCESO INTERACTIVO DE DATOS EN BLOODNET.

Aunque describir a nivel técnico la creación de esta Red Neuronal no es objetivo de esta tesis doctoral, a modo de resumen diremos que para su diseño se ha utilizado el lenguaje de programación Python, junto con las librerías de SciKit-Learn y Tensorflow, además de utilizar Google Colab como entorno de desarrollo. Su diseño al completo es de acceso público en el artículo publicado por nuestro equipo de trabajo (202).

En cuanto al balanceo de las muestras, hemos aplicado métodos basados en Funciones de Peso o *Sample Weight* en el entrenamiento del sistema. Mediante estos sistemas, dependiendo del número de muestras existentes en cada uno de los grupos, el peso de cada valor (alto o bajo) tendrá un valor distinto (aplicándose siempre de manera porcentual). No obstante, en esta tesis aplicamos sistemas aprendizaje automático supervisado con muestras etiquetadas de manera manual por un especialista y por ello no existen diferencias relevantes en el tamaño muestral de cada grupo, como se muestra en el escenario 3 de la **Tabla 7** (apartado 5.10).

## 5.10. Experimentos realizados

Una vez etiquetados los datos (**Tabla 3**), se han diseñado tres experimentos iniciales, al que hay que añadir un cuarto realizado previa aplicación de los criterios de exclusión (datos vírgenes), que hemos denominado experimento 3.5. Cada experimento corresponde a un escenario con una complejidad mayor. Los tres primeros escenarios se han llevado a cabo con y sin hemoglobinopatías estructurales, tema que será posteriormente discutido. Los diferentes escenarios que se han llevado a cabo durante el estudio, así como la distribución de las muestras en cada escenario están plasmados en la **Tabla 7**.

Escenario	Dificultad	Grupos diagnósticos incluidos (n)
1	Clasificación binaria	Sanos (863) y enfermos (2432)
2	Clasificación multiclase (3 grupos)	Sanos (863), anemias congénitas (1433) y anemias adquiridas (999).
3	Clasificación multiclase (5 grupos)	Sanos (863), Talasemias (635), hemoglobinopatías estructurales (798), anemias ferropénicas (495) y anemias por trastorno crónico (504).
3.5	Clasificación multiclase (5 grupos) previa limpieza de datos	Sanos (863), Talasemias (1.174), hemoglobinopatías estructurales (998), anemias ferropénicas (521) y anemias por trastorno crónico (504).

TABLA 7. ESCENARIOS DE CADA UNO DE LOS ENTRENAMIENTOS REALIZADOS.

## 5.11. Métodos de validación y explicabilidad del sistema de IA

La validación es una herramienta fundamental en IA para demostrar que el sistema está aprendiendo las características intrínsecas (en este caso del hemograma) para cada una de las enfermedades del entrenamiento y no únicamente memorizando resultado por resultado. Existen muchas formas de validación de sistemas y generalmente siempre se reserva una parte de la muestra para finalmente utilizarla en el modelo de validación que se decide aplicar.

En nuestra tesis doctoral, tras el etiquetado de las muestras separamos un 10% de la muestra para la validación de los ejercicios. Además, durante el proceso de entrenamiento, se usaron técnicas de validación cruzada como *K-Fold* y de regularización como *Dropout* y *L1/L2*.

De manera añadida, hemos recogido una segunda muestra de 250 hemogramas realizados entre 2019 y 2021 para llevar a cabo una validación externa basada en datos recientes no pertenecientes al conjunto inicial de datos, y así simular una puesta en producción del algoritmo inteligente.

Para finalizar, analizamos las mejores herramientas de explicabilidad de la IA aplicada en nuestro trabajo, decidiendo finalmente aplicar la herramienta de valores SHAP (Shapley Additive explanations) y los gráficos de Beeswarm, los cuales nos muestran el valor o peso de cada una de las variables a la hora de realizar el diagnóstico de cada enfermedad. Mostraremos los resultados de explicabilidad correspondientes a los escenarios dos y tres (203).

## 5.12. Métricas utilizadas y conceptos asociados

Para mostrar los resultados del trabajo utilizaremos la precisión, la exactitud y la sensibilidad de todos los experimentos previamente comentados. Además, mostraremos las matrices de confusión y las curvas ROC y PRC de la Red Neuronal creada para esta tesis doctoral (BloodNet).

**Valores perdidos:** Valores que, por cualquier razón, no aparecen en el lugar correspondiente de la variable en cuestión, generalmente debido a que corresponden a valores introducidos manualmente (por ejemplo, la edad o el sexo de un paciente). Tienen un valor importante en los estudios de IA ya que los sistemas inteligentes no pueden trabajar con valores perdidos y se deben solventar con el uso de diferentes técnicas.

**Accuracy o Exactitud:** Es la manera de saber de manera global qué porcentaje de las predicciones han sido correctas. Se calcula dividiendo los verdaderos positivos (VP) y verdaderos negativos (VN) entre todas las predicciones: VP y VN, así como falsos positivos (FP) y negativos (FN).

**Precisión o Valor predictivo positivo:** Muestra la dispersión de un test al repetirlo varias veces sobre una misma muestra. De manera práctica muestra qué posibilidad hay de que realmente tenga una enfermedad cuando el test salé positivo. Se calcula dividiendo los VP entre los VP + FP. En estadística tradicional se conoce como Valor predictivo positivo.

**Recall o Sensibilidad:** Su resultado nos muestra qué porcentaje de enfermos son clasificados como tal. Se calcula dividiendo los VP entre los VP+FN.

**Matriz de confusión:** Herramienta que nos permite evaluar visualmente como ha sido el desempeño o grado de acierto de un algoritmo de aprendizaje automático, donde las columnas representan la predicción y las filas representan la clasificación real.

**Curvas ROC (*Receiver Operating Characteristic*) y su AUC (*Area Under the Curve*):**

Corresponden a métodos estadísticos que podemos interpretar como una curva de capacidad diagnóstica y el área que ésta curva deja bajo ella. En ella se enfrentan la sensibilidad y el complementario de la especificidad (1-especificidad) y su mejor resultado está representado en el ángulo superior izquierdo de la gráfica, mientras que un test no discriminativo plasmará una línea diagonal que cruce la gráfica. A mayor puntuación, mayor capacidad de clasificar de un test, entendiéndose que estamos ante un buen test cuando los valores del AUC están entre 0,75 y 0,9; que el test es muy bueno cuando estamos entre 0,9 y 0,97 y ante un test excelente cuando estamos ante un test con AUC entre 0,97 y 1 (204).

**Curvas PRC (*Precision-Recall Curve*):** En este caso la gráfica nos muestra la relación entre la sensibilidad (Recall) y la precisión (valor predictivo positivo) con el fin de encontrar un punto lo más ajustado a la parte superior derecha de la curva. Este método estadístico es poco utilizado como método estadístico en sanidad, pero si se utiliza en ingeniería de Inteligencia Artificial.

### 5.13. Aspectos éticos

En el proceso y desarrollo del estudio no se ha tenido contacto con los pacientes y solamente se ha accedido a la información referida a los resultados hematimétricos obtenidos de una muestra de sangre, anonimizadas previamente. Al tratarse de un estudio retrospectivo donde no se han realizado intervenciones de ningún tipo sobre los pacientes y que, en cuanto al tratamiento de sus datos, se ha cumplido la Ley de protección de datos vigente, no se ha solicitado consentimiento informado de los pacientes para el uso de las muestras incluidas en el estudio. Este proyecto de investigación fue aprobado por el comité ético del Hospital Universitario Clínico San Carlos de Madrid en febrero del 2018.

## 6. Resultados

Hemos dividido los resultados en varios apartados. El primer apartado **[6.1]** refleja de manera ordenada las diferentes recogidas de datos y la distribución de las muestras etiquetadas para el entrenamiento y/o validación de nuestra Red Neuronal. El segundo apartado **[6.2]** corresponde a los resultados definitivos tras aplicar los filtros de limpieza y los criterios de exclusión. En el siguiente apartado **[6.3]** se expondrán los resultados obtenidos previa limpieza de datos, lo que hemos denominado como “datos vírgenes”. En ambos apartados los resultados se muestran medidos en exactitud, precisión y sensibilidad. Nuestra Red Neuronal (BloodNet) ha sido comparada con sistemas de ML tradicional como el Support Vector Machine (SVM), Random Forest (RF) y XGBoost, así como con una Red Neuronal publicada y validada por la comunidad científica (TabNet) y sus resultados aparecen en estos dos apartados. Estos resultados se representan en dos columnas: con y sin la participación de Hemoglobinopatías Estructurales (HE) por sus particularidades intrínsecas, que se discutirán posteriormente.

En el cuarto apartado **[6.4]** se ofrecen las matrices de confusión y curvas ROC de los resultados de BloodNet. En el quinto apartado **[6.5]** podemos ver los resultados ofrecidos por los sistemas de explicabilidad de BloodNet, donde explicaremos cuáles han sido las variables de mayor peso a la hora de clasificar las diferentes anemias. El sexto y último apartado **[6.6]** muestra los resultados obtenidos de nuestra puesta a punto o validación externa.

## 6.1. Distribución de las muestras utilizadas para el entrenamiento y validación del sistema.

Hemos recogido 111.144 hemogramas registrados en la base de datos del HCSC de Madrid entre los años 2011 y 2018, de los cuales se han etiquetado 4.060 con un diagnóstico de anemia o paciente sano. Con estos 4.060 hemogramas “vírgenes” hemos llevado a cabo el ejercicio correspondiente al escenario 3.5 únicamente.

Posteriormente se aplicaron los criterios de exclusión y los filtros de limpieza comentados en el [apartado 5.6](#) obteniéndose un total 3.295 muestras, utilizadas para el entrenamiento final del sistema en los escenarios 1, 2 y 3. Para la validación externa o puesta a punto del sistema hemos recogido hemogramas registrados desde enero del 2019 a marzo del 2021 y hemos etiquetado 250 muestras. Ambas distribuciones aparecen en la **Tabla 8**.

<b>Etiqueta</b>	<b>Muestras utilizadas para el entrenamiento (90%) y validación (10%).</b>	<b>Muestras utilizadas para la puesta a punto</b>
Sanos	863	50
Talasemias	675	71
Hemoglobinopatías estructurales	798	29
Anemias Ferropénicas	455	49
Anemias por Trastorno Crónico	504	51
<b>TOTAL</b>	<b>3295</b>	<b>250</b>

TABLA 8. DISTRIBUCIÓN DE LAS MUESTRAS RECOGIDAS PARA EL ENTRENAMIENTO, VALIDACIÓN Y PUESTA A PUNTO DE LA RED NEURONAL BLOODNET.

## 6.2 Resultados finales

Los resultados se exponen de menor a mayor complejidad, es decir en la **Tabla 9** se muestran los resultados del escenario 1 (sano vs enfermo) y posteriormente en la **Tabla 10** y **TABLA 11** se representan los resultados de los escenarios 2 y 3 respectivamente. En **negrita** se representa el mejor resultado obtenido en cada uno de los escenarios.

**TABLA 9. RESULTADOS OBTENIDOS EN EL ESCENARIO 1:**

**PACIENTES SANOS VS PACIENTES CON ANEMIA.**

SVM: SUPPORT VECTOR MACHINE; RF: RANDOM FOREST

Algoritmo	Con Hemoglobinopatías estructurales	Sin Hemoglobinopatías estructurales		
<b>Exactitud/Accuracy (%)</b>				
SVM	86,1	92,8		
RF	88,9	95,9		
XGboost	86,6	94,1		
BloodNet	<b>89,1</b>	<b>96,4</b>		
TabNet	85,7	92,9		
<b>Valor predictivo positivo/Precision (%)</b>				
	Sanos	Enfermos	Sanos	Enfermos
SVM	63	96	91	94
RF	<b>73</b>	92	<b>93</b>	96
XGboost	70	92	<b>93</b>	94
BloodNet	71	<b>97</b>	<b>93</b>	<b>98</b>
TabNet	69	93	90	94
<b>Sensibilidad /Recall (%)</b>				
	Sanos	Enfermos	Sanos	Enfermos
SVM	89	83	90	95
RF	78	<b>90</b>	93	<b>96</b>
XGboost	77	88	88	<b>96</b>
BloodNet	<b>93</b>	86	<b>96</b>	<b>96</b>
TabNet	82	87	89	95

**TABLA 10. RESULTADOS OBTENIDOS EN EL ESCENARIO 2:  
SANOS VS ANEMIAS ADQUIRIDAS VS ANEMIAS CONGÉNITAS.**

AC: ANEMIAS CONGÉNITAS; AA: ANEMIAS ADQUIRIDAS; SVM: SUPPORT VECTOR MACHINE; RF: RANDOM FOREST

Algoritmo	Con hemoglobinopatías estructurales	Sin hemoglobinopatías estructurales
<b>Exactitud/Accuracy (%)</b>		
SVM	74,6	87,3
RF	74,0	89,6
XGboost	74,1	87,9
BloodNet	<b>78,0</b>	<b>90,1</b>
TabNet	71,9	86,3

	Valor predictivo positivo/ Precision (%)			Valor predictivo positivo/ Precision (%)		
	Sanos	AC	AA	Sanos	AC	AA
SVM	66	86	74	88	80	88
RF	<b>77</b>	77	68	<b>93</b>	87	<b>89</b>
XGboost	72,00	79	<b>76</b>	88	<b>90</b>	87
BloodNet	75	<b>88</b>	74	<b>93</b>	86	88
TabNet	68	78	69	91	82	85

	Sensibilidad /Recall (%)			Sensibilidad /Recall (%)		
	Sanos	AC	AA	Sanos	AC	AA
SVM	87	59	87	89	86	83
RF	85	62	81	<b>94</b>	87	88
XGboost	<b>90</b>	65	80	91	81	90
BloodNet	88	<b>66</b>	<b>88</b>	90	<b>88</b>	<b>91</b>
TabNet	80	61	80	89	82	86

**TABLA 11. RESULTADOS OBTENIDOS EN EL ESCENARIO 3: SANOS VS TALASEMIAS VS HEMOGLOBINOPATÍAS ESTRUCTURALES VS ANEMIAS FERROPÉNICAS VS ANEMIAS POR TRASTORNO CRÓNICO.**

AC: ANEMIAS CONGÉNITAS; AA: ANEMIAS ADQUIRIDAS; SVM: SUPPORT VECTOR MACHINE; RF: RANDOM FOREST. TAL: TALASEMIAS; HE: HEMOGLOBINOPATÍAS ESTRUCTURALES; AF: ANEMIAS FERROPÉNICAS; ATC: ANEMIAS POR TRASTORNO CRÓNICO; SVM: SUPPORT VECTOR MACHINE; RF: RANDOM FOREST.

Algoritmo	Con hemoglobinopatías estructurales	Sin hemoglobinopatías estructurales
-----------	-------------------------------------	-------------------------------------

**Exactitud/Accuracy (%)**

SVM	66,9	82,4
RF	67,6	87,1
XGboost	68,0	84,3
BloodNet	<b>71,8</b>	<b>88,6</b>
TabNet	65,0	82,4

**Valor predictivo positivo/ precision (%)**

	Sanos	TAL	HE	AF	ATC	Sanos	TAL	AF	ATC
SVM	67	74	56	66	67	90	<b>89</b>	69	72
RF	69	74	53	64	<b>70</b>	<b>92</b>	87	75	<b>87</b>
XGboost	72	74	53	68	67	90	85	79	81
BloodNet	<b>76</b>	<b>76</b>	<b>69</b>	<b>73</b>	59	87	87	<b>81</b>	85
TabNet	62	69	34	60	69	88	85	71	79

**Sensibilidad/Recall (%)**

	Sanos	TAL	HE	AF	ATC	Sanos	TAL	AF	ATC
SVM	81	81	25	73	<b>85</b>	90	79	<b>71</b>	83
RF	86	79	30	71	77	<b>95</b>	84	<b>71</b>	<b>93</b>
XGboost	85	81	<b>34</b>	60	81	92	84	69	90
BloodNet	<b>93</b>	<b>89</b>	<b>34</b>	<b>75</b>	72	88	<b>86</b>	69	<b>93</b>
TabNet	83	77	22	56	62	93	82	67	79

## 6.3 Resultados con datos vírgenes (sin limpieza de datos)

En este apartado se presentan los resultados obtenidos previa limpieza de datos (4.060 hemogramas) y con hemoglobinopatías estructurales incluidas, al que hemos denominado escenario 3.5 ya que comparte la complejidad del escenario 3 (en el que sí se aplicaron los sistemas de limpieza de datos y criterios de exclusión).

**TABLA 12. RESULTADOS OBTENIDOS EN EL ESCENARIO 3.5: SANOS VS TALASEMIAS VS HEMOGLOBINOPATÍAS ESTRUCTURALES VS ANEMIAS FERROPÉNICAS VS ANEMIAS POR TRASTORNO CRÓNICO.**

*AC: ANEMIAS CONGÉNITAS; AA: ANEMIAS ADQUIRIDAS; SVM: SUPPORT VECTOR MACHINE; RF: RANDOM FOREST. TAL: TALASEMIAS; HE: HEMOGLOBINOPATÍAS ESTRUCTURALES; AF: ANEMIAS FERROPÉNICAS; ATC: ANEMIAS POR TRASTORNO CRÓNICO; SVM: SUPPORT VECTOR MACHINE; RF: RANDOM FOREST.*

Algoritmo	Con hemoglobinopatías estructurales				
<b>Exactitud/Accuracy (%)</b>					
SVM	54				
RF	<b>57</b>				
XGBoost	56				
BloodNet	56				
TabNet	54				
<b>Valor predictivo positivo/Precision (%)</b>					
	Sanos	TAL	HE	AF	ATC
SVM	54	76	<b>40</b>	61	49
RF	58	57	<b>40</b>	<b>69</b>	58
XGBoost	55	57	38	65	<b>65</b>
BloodNet	<b>60</b>	<b>84</b>	29	55	50
TabNet	56	54	30	56	54
<b>Sensibilidad/Recall (%)</b>					
	Sanos	TAL	HE	AF	ATC
SVM	84	34	26	74	82
RF	78	53	25	57	74
XGBoost	77	61	17	64	66
BloodNet	<b>87</b>	28	<b>32</b>	<b>75</b>	<b>88</b>
TabNet	76	<b>63</b>	5	61	61

## 6.4 Curvas ROC y matrices de confusión relativas a los experimentos con BloodNet

A continuación, mostramos los resultados de las matrices de confusión de los experimentos realizados con BloodNet, donde las columnas representan la etiqueta o clasificación elegida por el algoritmo inteligente denominado *predicted label*. Las filas representan la clasificación real o *true label* (etiquetado manual por hematólogo realizado en la primera fase del proyecto). En estas matrices lo que se muestra es el % de aciertos y errores que ha cometido el sistema para cada uno de los diagnósticos, es decir, cuantos verdaderos positivos, verdaderos negativos, falsos positivos y falsos negativos hay porcentualmente para cada uno de los diagnósticos de anemias.

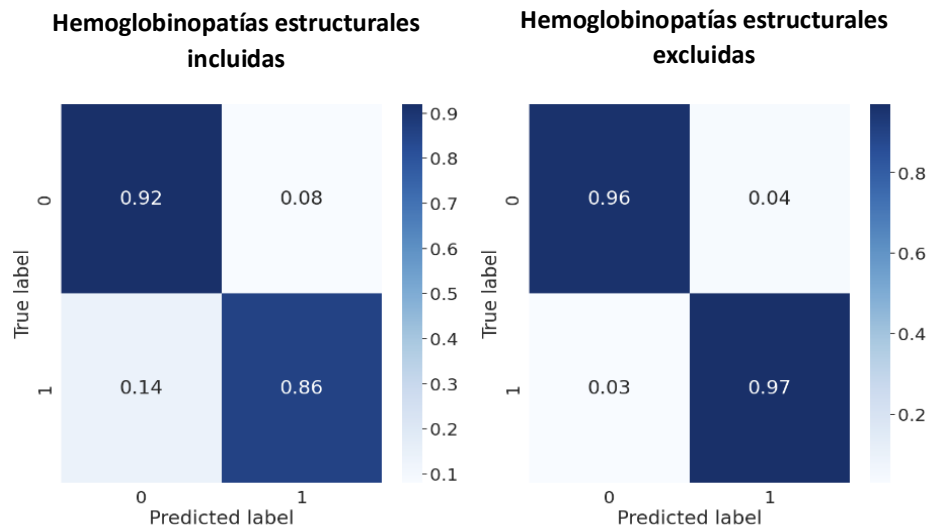
Se representan también las curvas ROC y su AUC de estos mismos experimentos, representándose en el eje X el inverso de la especificidad o tasa de falsos positivos (*false positive rate*) y en el eje Y la sensibilidad o tasa de verdaderos positivos (*true positive rate*).

Como comentamos en los métodos, en IA los resultados se suelen mostrar en Exactitud, Precisión o Recall y se suele utilizar otra curva denominada curva PRC que enfrenta precisión y el Recall.

Primero se representarán los resultados de las matrices de confusión y las curvas de los experimentos más sencillos (binarios), correspondientes al escenario 1 para posteriormente ir mostrando los experimentos de mayor complejidad progresivamente.

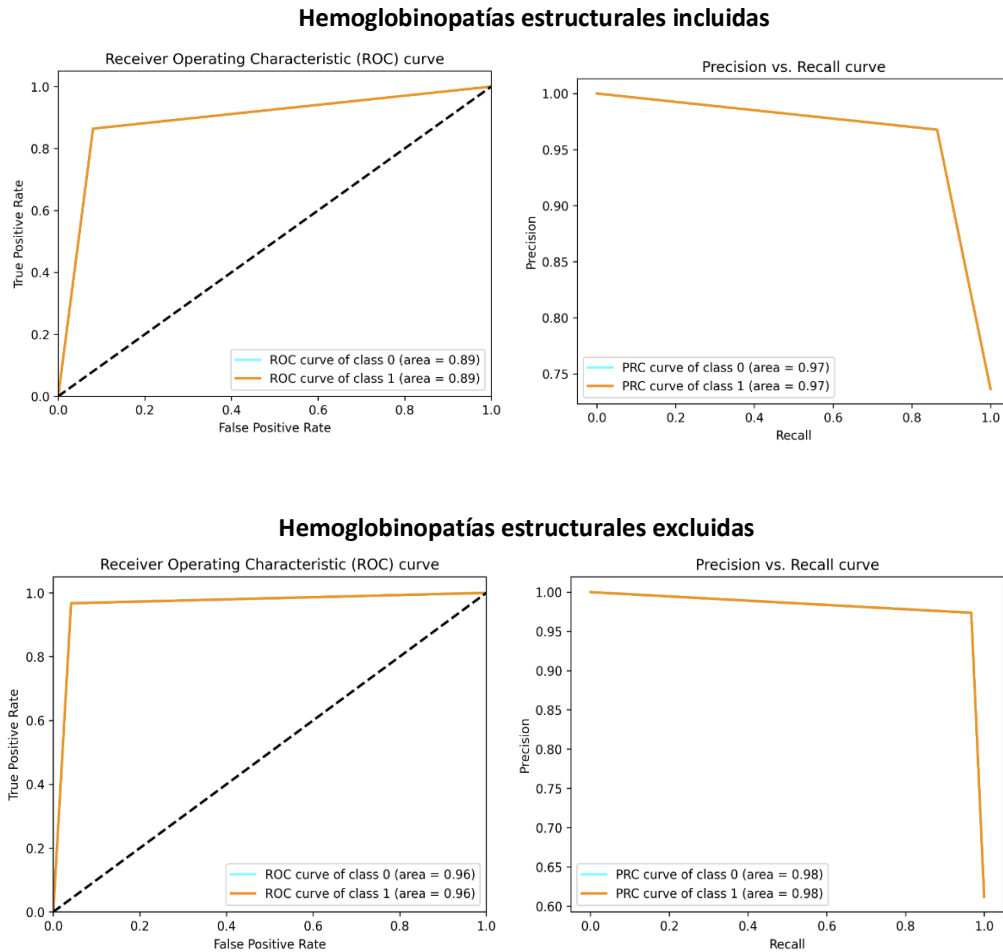
**FIGURA 19. MATRIZ DE CONFUSIÓN DE BLOODNET EN ESCENARIO 1:  
ENFERMOS VS SANOS.**

CLASE 0: PACIENTE SANO; CLASE 1: PACIENTE CON ANEMIA.



**FIGURA 20. CURVAS ROC (IZQUIERDA) Y PRC (DERECHA) DE BLOODNET EN ESCENARIO 1:  
ENFERMOS VS SANOS**

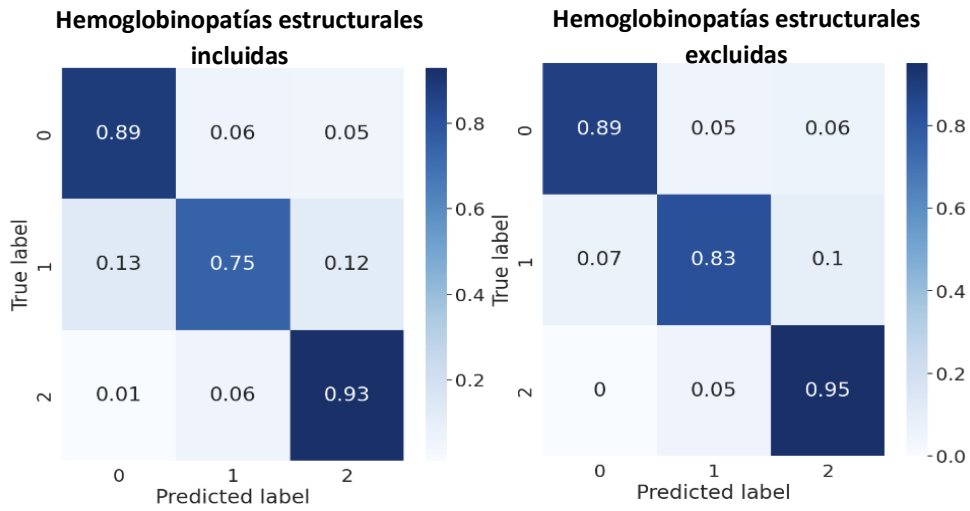
CLASE 0: PACIENTE SANO; CLASE 1: PACIENTE CON ANEMIA.



**FIGURA 21. MATRIZ DE CONFUSIÓN DE BLOODNET EN ESCENARIO 2:**

**SANOS VS ANEMIAS ADQUIRIDAS VS ANEMIAS CONGÉNITAS.**

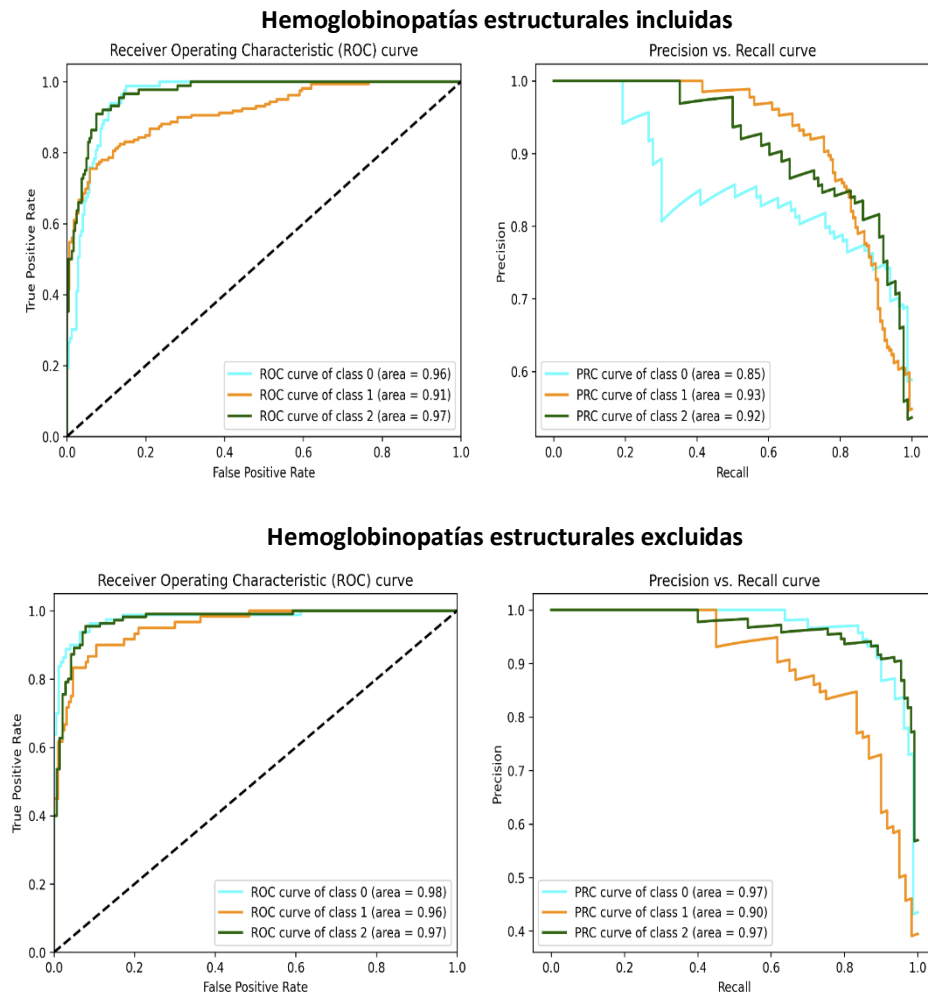
*CLASE 0: SANOS; CLASE 1: ANEMIAS CONGÉNITAS; CLASE 2: ANEMIAS ADQUIRIDAS*



**FIGURA 22. CURVAS ROC (IZQUIERDA) Y PRC (DERECHA) DE BLOODNET EN ESCENARIO 2:**

**SANOS VS ANEMIAS ADQUIRIDAS VS ANEMIAS CONGÉNITAS.**

*CLASE 0: SANOS; CLASE 1: ANEMIAS CONGÉNITAS; CLASE 2: ANEMIAS ADQUIRIDAS.*

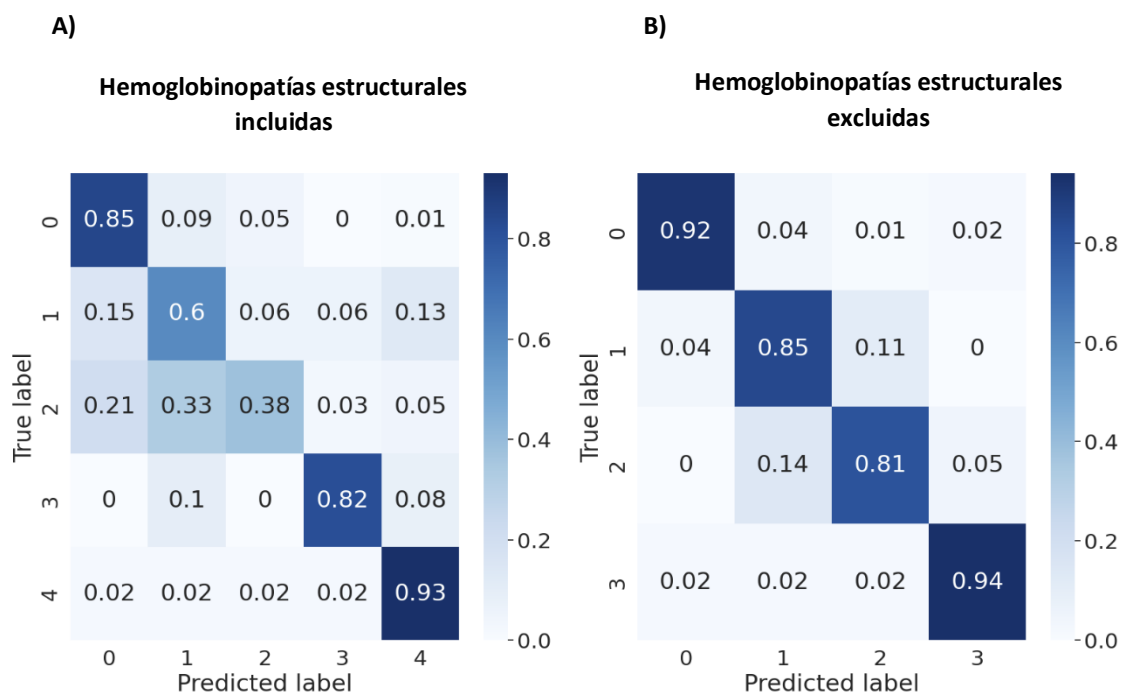


**FIGURA 23. MATRIZ DE CONFUSIÓN DE BLOODNET EN ESCENARIO 3:**

**SANOS VS TALASEMIAS VS HEMOGLOBINOPATÍAS ESTRUCTURALES VS ANEMIAS FERROPÉNICAS VS ANEMIAS POR TRASTORNO CRÓNICO.**

**A) CLASE 0: PACIENTES SANOS, CLASE 1: TALASEMIAS, CLASE 2: HEMOGLOBINOPATÍAS ESTRUCTURALES, CLASE 3: ANEMIA FERROPÉNICA, CLASE 4: ANEMIA POR TRASTORNO CRÓNICO.**

**B) CLASE 0: PACIENTES SANOS, CLASE 1: TALASEMIAS, CLASE 2: ANEMIA FERROPÉNICA, CLASE 3: ANEMIA POR TRASTORNO CRÓNICO.**

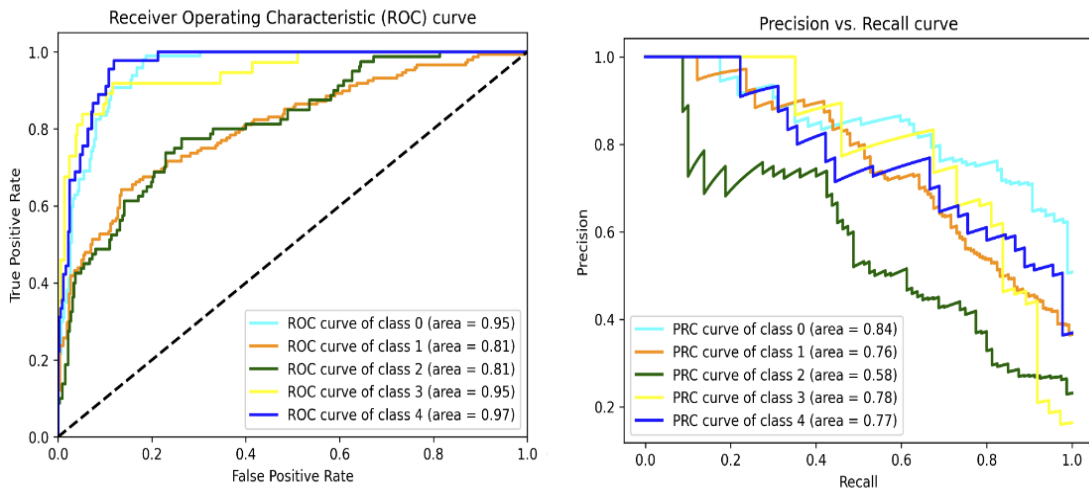


**FIGURA 24. CURVAS ROC (IZQUIERDA) Y PRC (DERECHA) DE BLOODNET EN ESCENARIO 3:  
SANOS VS TALASEMIAS VS HEMOGLOBINOPATÍAS ESTRUCTURALES VS ANEMIAS FERROPÉNICAS  
VS ANEMIAS POR TRASTORNO CRÓNICO.**

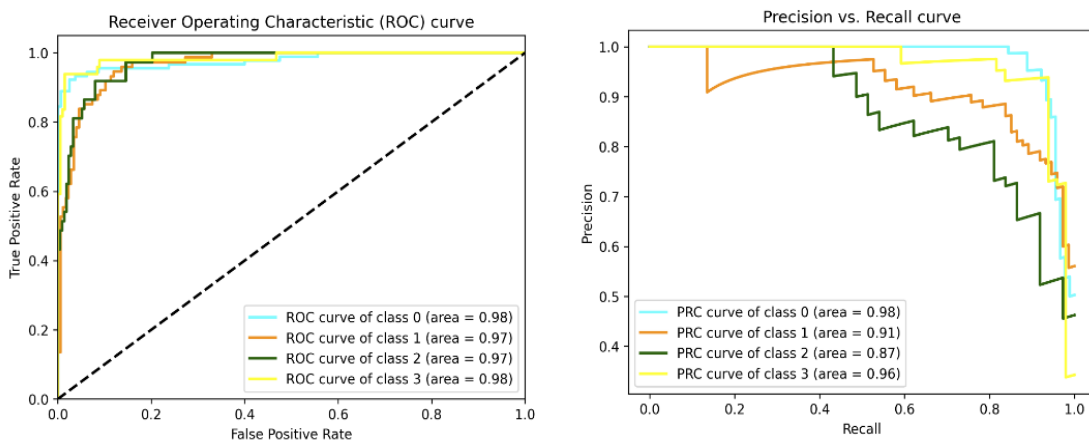
**A) CLASE 0: PACIENTES SANOS, CLASE 1: TALASEMIAS, CLASE 2: HEMOGLOBINOPATÍAS ESTRUCTURALES, CLASE 3: ANEMIA FERROPÉNICA, CLASE 4: ANEMIA POR TRASTORNO CRÓNICO.**

**B) CLASE 0: PACIENTES SANOS, CLASE 1: TALASEMIAS, CLASE 2: ANEMIA FERROPÉNICA, CLASE 3: ANEMIA POR TRASTORNO CRÓNICO.**

**A) Hemoglobinopatías estructurales incluidas**



**B) Hemoglobinopatías estructurales excluidas**



## 6.5 Resultados de explicabilidad utilizados sobre BloodNet.

A continuación, se presentan los resultados de los gráficos de explicabilidad aplicados para cada uno de los diagnósticos de anemias en los escenarios 2 y 3.

Los valores SHAP y las gráficas de BEESWARM nos muestran de un solo vistazo el peso de cada una de las variables utilizadas en cada uno de los experimentos (escenarios). A la izquierda veremos un gráfico de barras (*Bar Plot*) en el que se suman los pesos o la importancia de cada variable en cada escenario. Arriba se colocarán las variables más importantes a la hora de diferenciar las anemias de dicho escenario.

En el lado derecho de cada Figura se representa la importancia de cada variable para cada diagnóstico concreto de anemia (talasemias, por ejemplo) dentro de un escenario concreto y también se colocarán arriba las variables de mayor interés y abajo las de menor interés.

En estos gráficos, los puntos azules corresponden a valores bajos de cada variable (Hemoglobina, por ejemplo) y los rojos a valores altos. Por otro lado, los puntos posicionados a la izquierda de la línea divisoria central corresponden a valores de bajo interés en el diagnóstico en cuestión. Asimismo, los puntos colocados a la derecha se traducen en que los valores altos o bajos (puntos rojos o azules respectivamente) de dicha variable en cuestión han sido de gran interés a la hora de decidir el diagnóstico por nuestra Red Neuronal.

Comentaremos cada uno de los resultados para que sea sencillo de entender.

### 6.5.1 Gráficos de explicabilidad en el escenario 2: Pacientes sanos

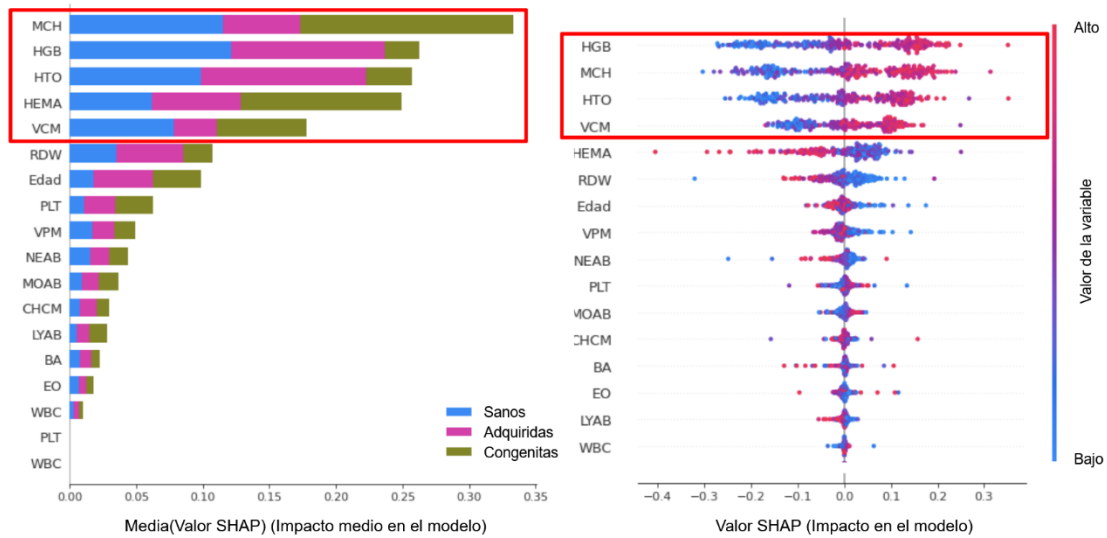


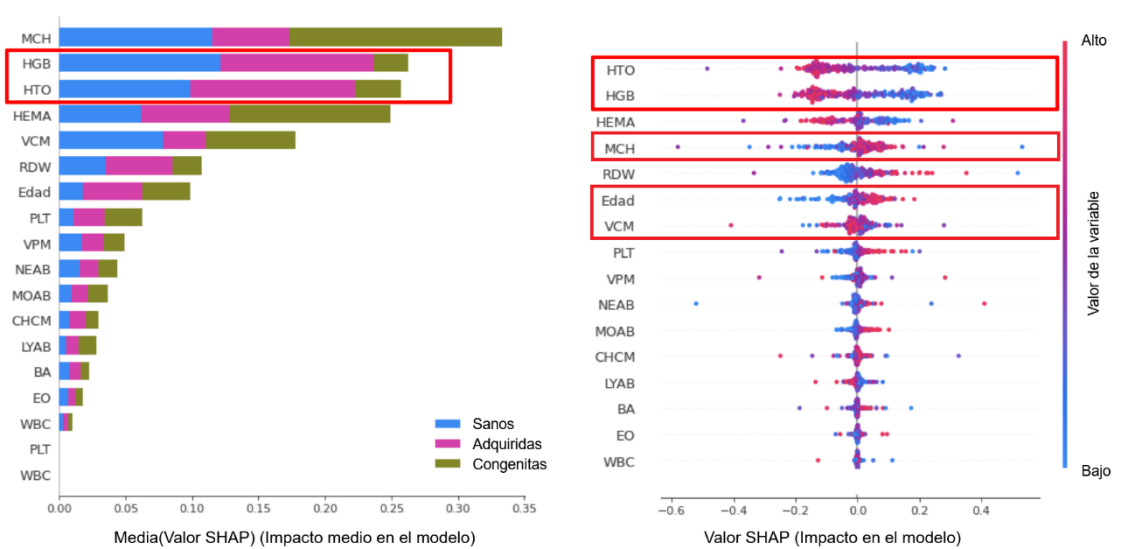
FIGURA 25. GRÁFICOS DE EXPLICABILIDAD PARA PACIENTES SANOS (ESCENARIO 2) MEDIANTE BLOODNET.

Lo primero que debemos tener en cuenta es que en el escenario 2 estamos diferenciando entre pacientes sanos, pacientes con anemias adquiridas y anemias congénitas. A la izquierda de la **Figura 25** vemos como el algoritmo ha seleccionado las variables de HCM, HGB, HTO y HEMA como las cuatro de mayor interés para el diagnóstico de anemias en el escenario 2 (de manera global).

A la derecha de la gráfica vemos que, para la clasificación de pacientes sanos dentro del escenario 2, BloodNet le ha dado mayor peso a valores altos o normales de HGB, HCH, HTO y VCM (representados como puntos rojos y morados en la parte central y derecha de la gráfica). Sin embargo, los valores bajos de estas mismas variables (representados en azul) quedan a la izquierda de esta gráfica y marcan un bajo interés para nuestro sistema inteligente.

## 6.5.2 Gráficos de explicabilidad en el escenario 2: Pacientes con anemias adquiridas

A la derecha vemos que para clasificar en el escenario 2 a los pacientes con anemias adquiridas, BloodNet se ha fijado en valores bajos en las variables HGB y HTO (de color azul y posicionados a la derecha de la gráfica), además de fijarse en los valores normales o ligeramente elevados de las variables HCM y Edad (**Figura 26**).



**FIGURA 26. GRÁFICOS DE EXPLICABILIDAD PARA PACIENTES CON ANEMIAS ADQUIRIDAS (ESCENARIO 2) MEDIANTE BLOODNET.**

6.5.3 Gráficos de explicabilidad en el escenario 2: Pacientes con anemias congénitas (talasemias).

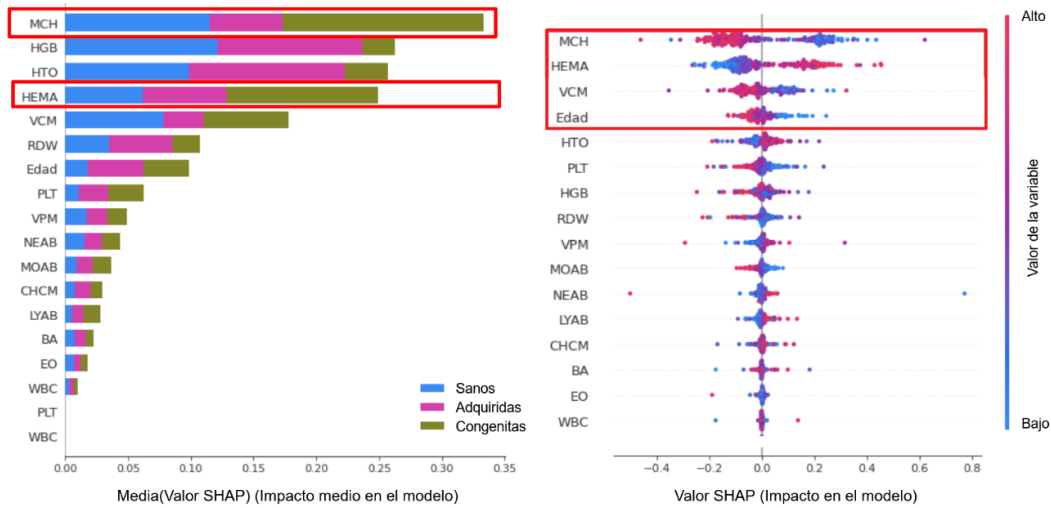
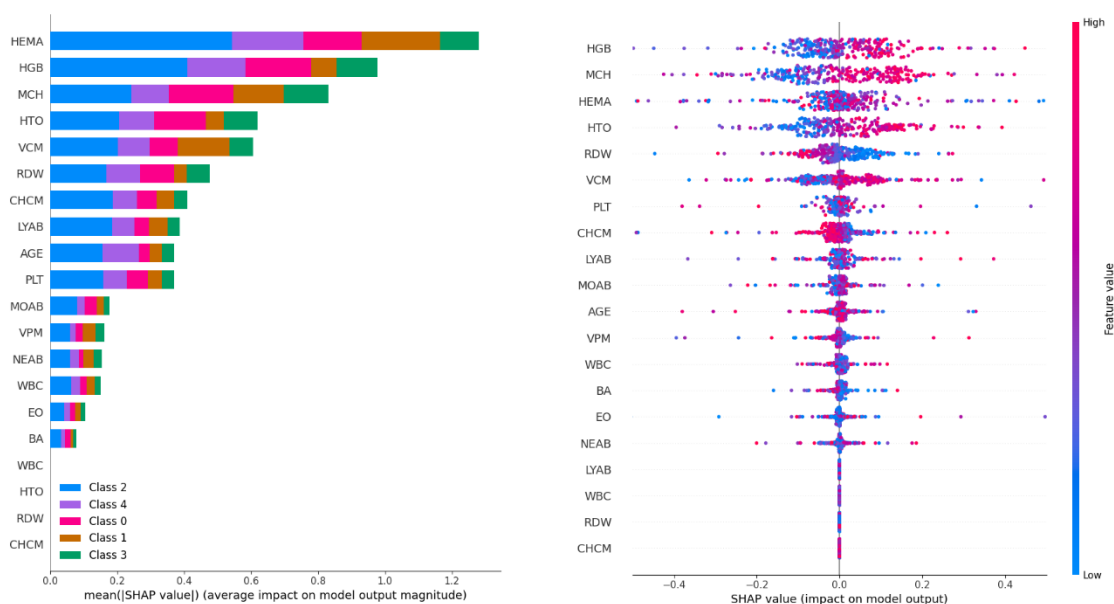


FIGURA 27. GRÁFICOS DE EXPLICABILIDAD PARA PACIENTES CON TALASEMIAS (ESCENARIO 2) MEDIANTE

**BLOODNET.**

En este caso de las anemias congénitas, en las que se excluyeron las HE, vemos a la derecha como el diagnóstico de Talasemias por BloodNet está basado en valores de las variables HCM bajos (valores azules a la derecha), valores de HEMA elevados (puntos rojos a la derecha) y tanto valores bajos o discretamente disminuidos de las variables de VCM y edad (Figura 27).

#### 6.5.4 Gráficos de explicabilidad en el escenario 3: pacientes sanos



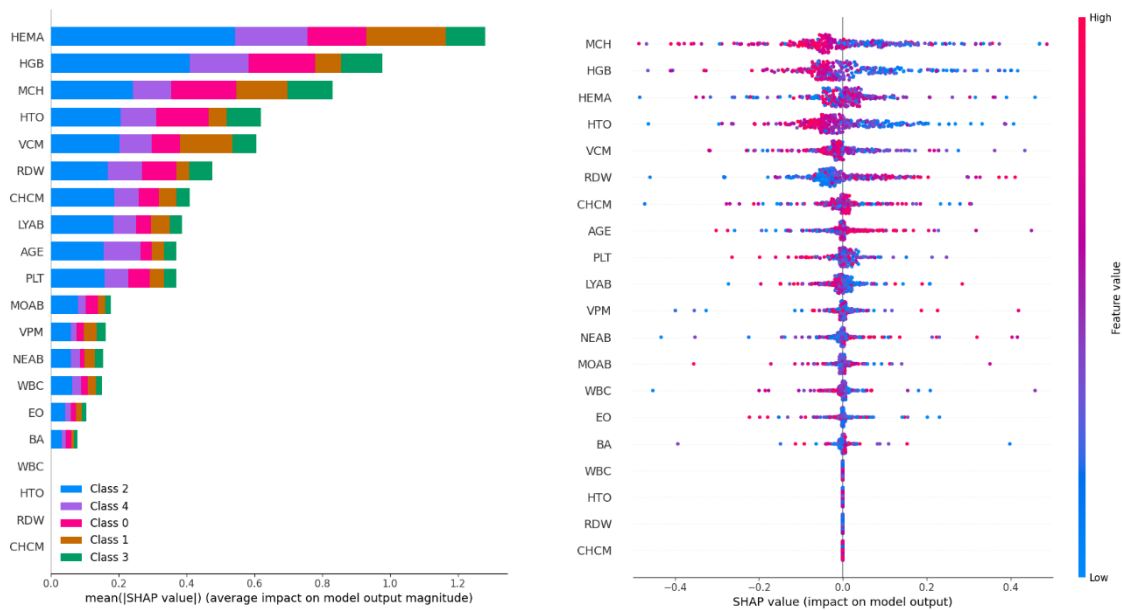
**FIGURA 28. GRÁFICOS DE EXPLICABILIDAD PARA PACIENTES SANOS (ESCENARIO 3) MEDIANTE BLOODNET.**

*CLASE 0: PACIENTES SANOS, CLASE 1: TALASEMIAS, CLASE 2: HEMOGLOBINOPATÍAS ESTRUCTURALES, CLASE 3: ANEMIA FERROPÉNICA, CLASE 4: ANEMIA POR TRASTORNO CRÓNICO.*

En este caso nos encontramos en el escenario 3, en el que diferenciamos entre las diferentes 4 anemias y los pacientes sanos. A la izquierda vemos como los valores HEMA, HGB, HCM, HTO y VCM son los de mayor interés para BloodNet en todo el escenario 3 (Figura 28).

A la derecha, valorando la identificación de pacientes sanos vemos como volvemos a tener valores elevados o normales de HGB y HCM, así como valores normales de HEMA como las variables de mayor interés.

### 6.5.5 Gráficos de explicabilidad en el escenario 3: Anemias ferropénicas



**FIGURA 29. GRÁFICOS DE EXPLICABILIDAD PARA PACIENTES CON ANEMIA FERROPÉNICA (ESCENARIO 3) MEDIANTE BLOODNET.**

*CLASE 0: PACIENTES SANOS, CLASE 1: TALASEMIAS, CLASE 2: HEMOGLOBINOPATÍAS ESTRUCTURALES, CLASE 3: ANEMIA FERROPÉNICA, CLASE 4: ANEMIA POR TRASTORNO CRÓNICO.*

En la **Figura 29** podemos observar como el algoritmo inteligente BloodNet se ha fijado en los valores bajos de HCM, HBG, HTO (puntos azules a la derecha), así como en los valores altos de RDW para alcanzar una buena clasificación de anemias ferropénicas en el escenario 3.

## 6.5.6 Gráficos de explicabilidad en el escenario 3: Anemias por trastorno crónico

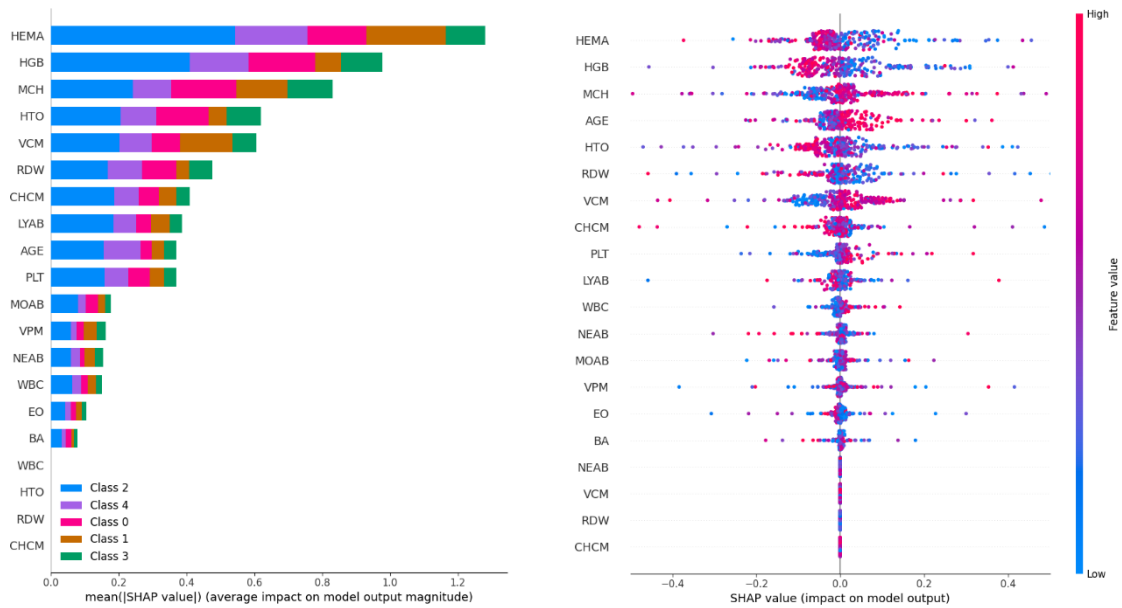


FIGURA 30. GRÁFICOS DE EXPLICABILIDAD PARA PACIENTES CON ANEMIA POR TRASTORNO CRÓNICO (ESCENARIO 3)

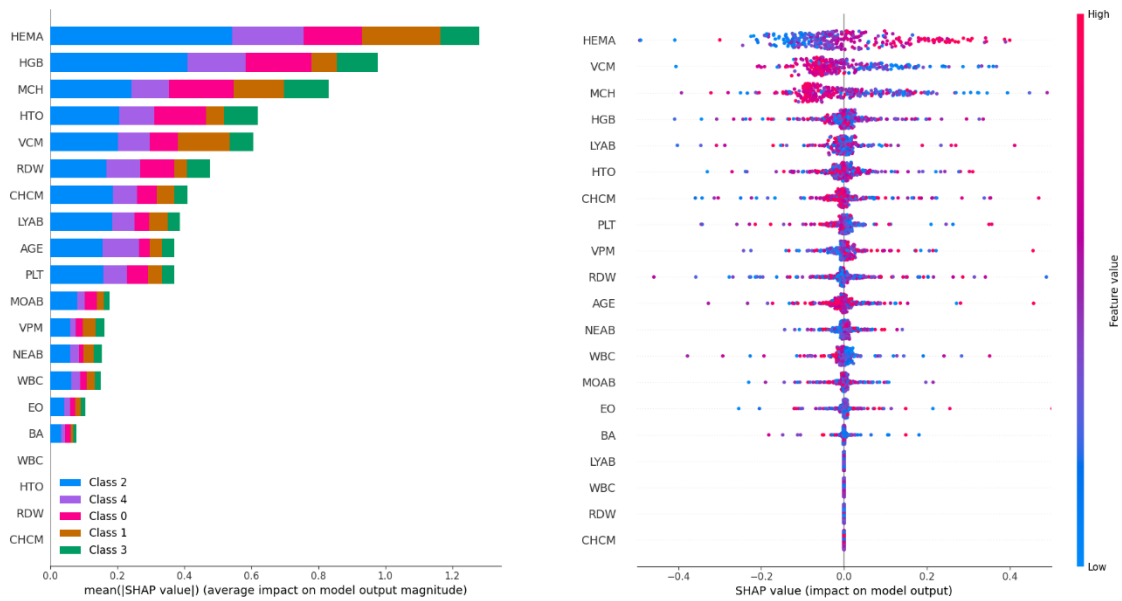
MEDIANTE BLOODNET.

CLASE 0: PACIENTES SANOS, CLASE 1: TALASEMIAS, CLASE 2: HEMOGLOBINOPATÍAS ESTRUCTURALES, CLASE 3: ANEMIA FERROPÉNICA,

CLASE 4: ANEMIA POR TRASTORNO CRÓNICO.

En el caso de las anemias de trastorno crónico, a la derecha vemos como gana interés la variable Edad, que se coloca en cuarta posición. Esto, junto a los valores bajos de HGB y HEMA, los valores medios de HCM y un valor de VCM medio/alto llevan al algoritmo a conseguir el diagnóstico (Figura 30).

### 6.5.7 Gráficos de explicabilidad en el escenario 3: Talasemias



**FIGURA 31. GRÁFICOS DE EXPLICABILIDAD PARA PACIENTES CON ANEMIA POR TALASEMIAS (ESCENARIO 3)**

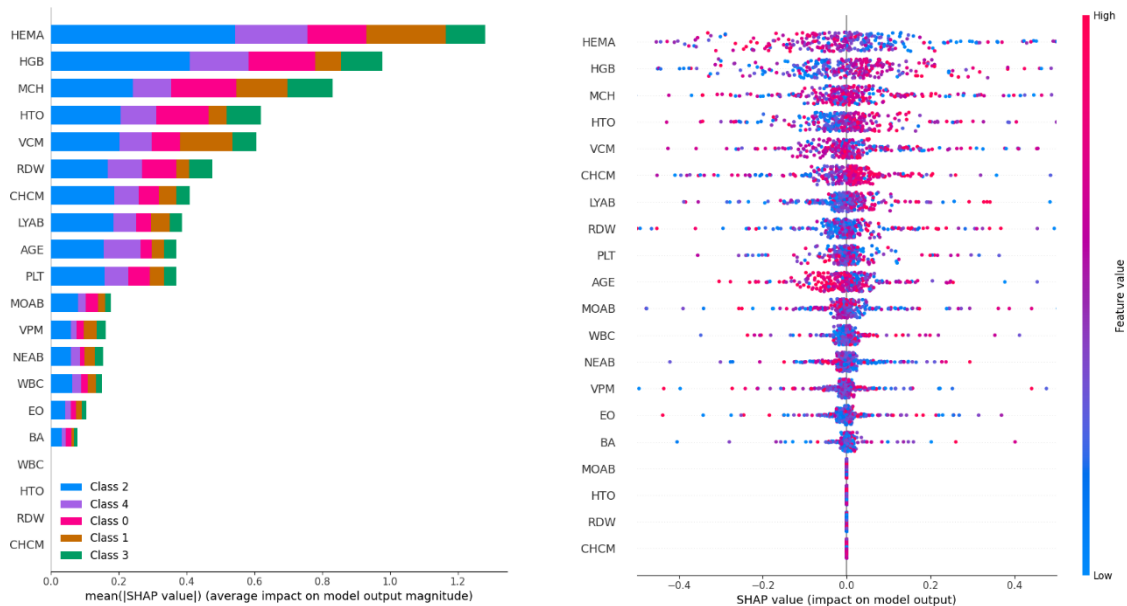
**MEDIANTE BLOODNET.**

*CLASE 0: PACIENTES SANOS, CLASE 1: TALASEMIAS, CLASE 2: HEMOGLOBINOPATÍAS ESTRUCTURALES, CLASE 3: ANEMIA FERROPÉNICA,*

*CLASE 4: ANEMIA POR TRASTORNO CRÓNICO.*

En la **Figura 31** vemos que para el diagnóstico de Talasemias, BloodNet se ha guiado de valores altos de HEMA (puntos rojos a la derecha), así como de valores bajos de VCM y HCM (puntos azules a la derecha).

## 6.5.8 Gráficos de explicabilidad en el escenario 3: Hemoglobinopatías estructurales



**FIGURA 32. GRÁFICOS DE EXPLICABILIDAD PARA PACIENTES CON ANEMIA POR HEMOGLOBINOPATÍAS ESTRUCTURALES (ESCENARIO 3) MEDIANTE BLOODNET.**

*CLASE 0: PACIENTES SANOS, CLASE 1: TALASEMIAS, CLASE 2: HEMOGLOBINOPATÍAS ESTRUCTURALES, CLASE 3: ANEMIA FERROPÉNICA, CLASE 4: ANEMIA POR TRASTORNO CRÓNICO.*

En el caso de las hemoglobinopatías estructurales representadas en la **Figura 32** podemos ver cómo es complicado identificar un patrón claro de clasificación ya que tenemos una gran heterogeneidad en la dispersión de los puntos (existen puntos azules y rojos cruzados en todas las zonas del gráfico). Esto se traduce en una baja precisión de diagnóstico de esta entidad por BloodNet para las hemoglobinopatías estructurales.

## 6.6 Resultados de la puesta a punto o validación externa de BloodNet.

A continuación, mostramos los resultados obtenidos de la puesta a punto mediante el análisis de 250 muestras recogidas tras la finalización del trabajo, cuya distribución por diagnósticos aparece en la **Tabla 8** del apartado **6.1**. Se presentan los resultados de un escenario 3 sin HE incluidas, en el que diagnostican pacientes sanos, talasemias, anemias ferropénicas y anemias por trastorno crónico. Los resultados medidos en sensibilidad, precisión y exactitud se muestran en la **Tabla 13** y la matriz de confusión correspondiente a este ejercicio se muestra en la **Figura 33**.

Algoritmo BloodNet	Sanos	Talasemias	Anemias Ferropénicas	Anemias Trastorno Crónico
Precisión	1	0,9	0,82	0,92
Sensibilidad	0,94	0,9	0,86	0,92
Exactitud	0,9			

TABLA 13. RESULTADOS OBTENIDOS DE LA PUESTA A PUNTO O VALIDACIÓN EXTERNA DE BLOODNET

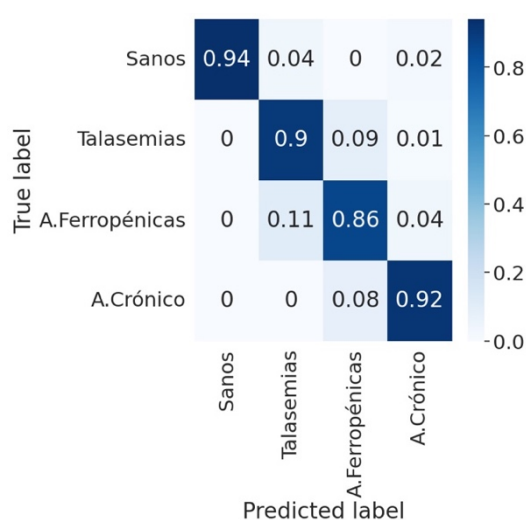


FIGURA 33. MATRIZ DE CONFUSIÓN DE LA VALIDACIÓN EXTERNA O PUESTA A PUNTO DE BLOODNET

# 7. Discusión

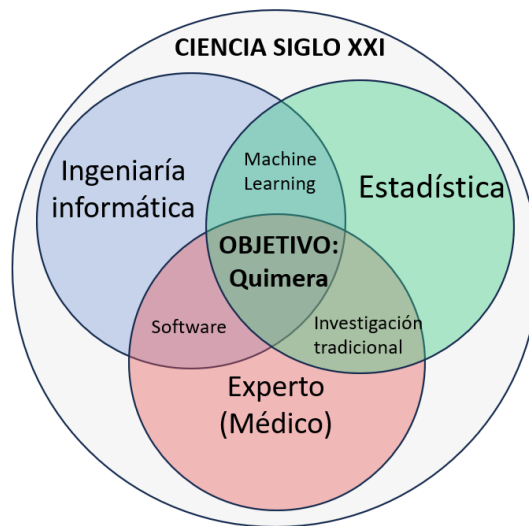
## 7.1 El valor de la Inteligencia Artificial en Medicina. Evaluación de resultados

La medicina, como cabeza representante de las ciencias de la salud, debe crear coalición con las ciencias computacionales para mantenerse actualizada durante el avance del siglo XXI para continuar ofreciendo los mejores diagnósticos y tratamientos a los pacientes.

Uno de los mayores obstáculos para poder iniciar un proyecto de Inteligencia Artificial (IA) en medicina es encontrar dos equipos de trabajo de diferentes ámbitos con un objetivo común. Es decir, un equipo de ingeniería informática que quiera aprender medicina para entender qué está buscando en el diseño de sus algoritmos inteligentes y un equipo médico que desee investigar los aportes de la IA en el diagnóstico de enfermedades, para lo cual deberá entender las bases de esta ciencia. Atrás deben quedar las ideas de que la Inteligencia Artificial llega para sustituir al médico para dejar paso a las que desean contribuir en la búsqueda de una medicina personalizada y de alta precisión.

Bajo nuestro punto de vista y después de lo aprendido en estos años de trabajo, el futuro médico deberá tener nociones básicas de IA si quiere entender las investigaciones en las que se aplique, cada vez más elevadas (**Figura 1**). Será similar al inicio del manejo de ordenadores en 1985 o al aprendizaje de los sistemas estadísticos ya normalizados a día de hoy en investigación (14,205,206).

En los próximos años vamos a vivir la entrada al mercado de jóvenes licenciados provenientes de carreras donde estos conocimientos ya se entrecruzan como ingeniería biomédica, biotecnología o bioingeniería. Sin embargo, hasta que se asienten y proliferen debemos formar equipos multidisciplinares que nos permitan compartir de manera cruzada los conocimientos necesarios para continuar investigando en cuáles son los mejores sistemas de análisis de datos sanitarios de los que disponemos, que cada vez es mayor. La evolución en estos equipos multidisciplinares de investigación entre el siglo XX y siglo XXI está representada en la **Figura 34**



**FIGURA 34. REPRESENTACIÓN DE LA MULTIDISCIPLINARIEDAD EN LA CIENCIA DEL SIGLO XXI.**

En esta tesis, iniciada en 2018, cuando la IA aún no era tendencia (**Figura 2**), hemos demostrado que la coalición de equipos multidisciplinares de instituciones y ciudades diferentes es posible y, en este caso, nos ha permitido diseñar un sistema inteligente capaz de separar pacientes sanos y enfermos con una Exactitud (*Accuracy*) de hasta 96% en su mejor escenario y que incluso en su peor escenario ha mantenido una Exactitud del 88,6% (**Tabla 14**).

Algoritmo	Accuracy			
	Escenario 1 (binario)	Escenario 2 (3 clases)	Escenario 3 (4 clases)	Escenario 3.5 (datos vírgenes)
SVM	0,928	0,873	0,824	0,54
Random Forest	0,959	0,896	0,871	0,57
XGBoost	0,941	0,879	0,843	0,56
BloodNet	<b>0,964</b>	<b>0,901</b>	<b>0,886</b>	0,56
TabNet	0,929	0,863	0,824	0,54

TABLA 14. RESUMEN DE LA EXACTITUD/ACCURACY DE LOS EXPERIMENTOS REALIZADOS. SE HAN EXCLUIDO LAS

HEMOGLOBINOPATÍAS ESTRUCTURALES, EXCEPTO EN EL ESCENARIO 3.5

Los resultados obtenidos en los diferentes experimentos realizados son de mayor calidad en los escenarios iniciales de menor complejidad, donde los sistemas inteligentes han llevado a cabo una clasificación binaria entre sujetos sanos y pacientes con anemias, llegando a obtenerse resultados con nuestra Red Neuronal que superan una exactitud del 96%. A medida que la complejidad del trabajo se ha aumentado, con la diferenciación de las distintas causas de anemia, los resultados obtenidos han disminuido, pero han mantenido exactitudes del 88,6 % en su escenario más complejo (escenario 3).

Podemos ver como los resultados del resto de sistemas operativos inteligentes utilizados no se quedan atrás y muestran resultados competitivos. De hecho, podemos ver cómo los sistemas de ML tradicional (Random Forest, sobre todo) se acercan mucho a los resultados obtenidos con nuestra Red Neuronal (BloodNet) e incluso superan a la Red Neuronal de acceso público (TabNet). Esto se debe principalmente a que los

sistemas de ML tradicional son capaces de obtener resultados muy competentes cuando se trabaja con datos tabulados (números) y mejorar sus resultados con las Redes Neuronales se puede convertir en un objetivo complicado de cumplir. Es por ello que la mayoría de los sistemas de IA basados en Redes Neuronales están enfocados a mejorar otros campos que le quedan grandes a los sistemas de ML tradicional: reconocimiento de imagen y lenguaje natural, como se demuestra en publicaciones recientes de IA-hematología (105,124,207).

Sin embargo, esto no debe frenar nuestra intención de continuar mejorando los sistemas de diagnóstico basados en el análisis de datos numéricos. Tengamos presente que la mejora en la detección de un 1% en una enfermedad de alta prevalencia puede conseguir mejoras en el diagnóstico de miles de personas. En este sentido, recordemos que, según los organismos internacionales como OMS y publicaciones recientes, la prevalencia de la anemia a nivel mundial se encuentra entre un 24,8 % y 32,9 % (1,2).

Por ello creemos que, a la luz de nuestros resultados, la mejora que supone aplicar BloodNet frente a los sistemas inteligentes tradicionales ya existentes, cumple el objetivo marcado de mejora de diagnóstico en anemias basado en Redes Neuronales.

Para explicar de manera más desgranada los resultados de nuestra Red Neuronal hemos mostrado las matrices de confusión, las curvas ROC y las curvas PRC explicándose su utilidad en el apartado de resultados (**Figura 19, Figura 20, Figura 21, Figura 22, Figura 23 y Figura 24**). Estas matrices y curvas las hemos replicado únicamente para mostrar los resultados de BloodNet ya que el resto de algoritmos inteligentes (tanto tradicionales como el de TabNet) ya están validadas a nivel científico y no es objeto de esta tesis doctoral justificar su funcionamiento.

Si analizamos cuáles han sido las anemias que mejor ha conseguido diagnosticar nuestra Red Neuronal en el escenario 3, vemos como las Anemias de Trastorno Crónico junto con los Pacientes Sanos han alcanzado un AUC de 0,98 mientras que las Talasemias y las Anemias Ferropénicas están por detrás con AUC de 0,97 (apartado B de la **Figura 24**). Como se explica en el [apartado 5.12](#), los resultados de AUC comprendidos entre 0,97-1, se consideran como excelentes en la literatura científica (204).

Estas diferencias probablemente están basadas en la heterogeneidad que pueden presentar los hemogramas de pacientes con Anemia Ferropénica y los diferentes tipos de Talasemia, frente a los valores más homogéneos entre sí que muestran los pacientes con Anemias por Trastorno Crónico y Pacientes Sanos. Estos datos más homogéneos entre si hacen que el aprendizaje de la Red Neuronal sea más eficaz, pero esto se discutirá en el apartado de explicabilidad.

## 7.2 El valor de la calidad y cantidad de datos

Una posible crítica de este tipo de trabajos es la de validar el papel del médico. Este papel lo debemos entender desde varios puntos de vista: a) el etiquetado de las muestras debe ser riguroso y de alta calidad, b) la elaboración del diseño del algoritmo inteligente con conocimiento específico mejora la capacidad de aprendizaje del sistema y c) la limpieza de datos aplicando los criterios de exclusión seleccionados previamente por el especialista mejoran la capacidad del sistema inteligente. En resumen, la calidad de los datos etiquetados previo análisis de los mismos es indispensable para un correcto aprendizaje del sistema. En una publicación reciente del 2023, *Obstfeld et al.* mencionan el sesgo de los datos o el error a la hora de etiquetarlos como uno de los puntos débiles en la IA en hematología a día de hoy (112).

Nosotros hemos confirmado en esta tesis que la depuración de la muestra, es decir, la selección de los criterios de inclusión y exclusión mejora el aprendizaje del sistema, demostrando esto en la drástica diferencia de los resultados obtenidos entre el escenario 3 frente al escenario 3.5 (**Tabla 11**,

**Tabla 12 y Tabla 14**).

La IA aprende a partir de los ejemplos que le aportamos: a mayor cantidad de datos, mayor será su capacidad de discriminación ante los pequeños detalles que pueden existir entre las diferentes patologías (203). Antes de iniciar un trabajo donde se pretendan aplicar Redes Neuronales se debe valorar la capacidad de acceso a la cantidad de datos necesarios para llevar a cabo un aprendizaje automático. En el caso de que la cantidad de datos sea baja, nos debemos plantear un algoritmo más sencillo, donde entran a jugar un papel importante los criterios de inclusión y exclusión.

En esta tesis doctoral, aunque partimos de un número elevado de muestras con 4.060 hemogramas (colocándonos entre los trabajos con mayor muestra en IA y hematología), seguía siendo un número limitado para agregar características específicas que aportaran gran complejidad al problema. En este sentido, por ejemplo, excluimos a los pacientes que se hubieran transfundido concentrados de hematíes en los últimos tres meses del hemograma en cuestión, ya que el sistema no tenía suficientes datos como para distinguir entre los pacientes transfundidos con anemia ferropénica y los pacientes sanos.

Algo parecido ha ocurrido con los pacientes con hemoglobinopatías estructurales. En este caso, debido a las características intrínsecas de la patología, la máquina inteligente no ha alcanzado resultados óptimos, consiguiéndose exactitudes del 78% y 71,2% en el escenario 2 y 3 respectivamente (**Tabla 10** y **TABLA 11**) y curvas ROC con un AUC entre 0,91 y 0,81 en los escenarios 2 y 3 respectivamente (**Figura 23** y **Figura 24**). Estos resultados nos han llevado a hacer siempre dos análisis en cada escenario: con y sin hemoglobinopatías estructurales.

Podemos justificar esta dificultad en el aprendizaje por:

1. Existe un abanico de altísima variabilidad entre los diferentes subtipos de hemoglobinopatías estructurales incluidas en el estudio (**Tabla 5**), siendo en la mayoría de las ocasiones pacientes portadores heterocigotos de alteraciones genéticas que no se reflejan en el hemograma.
2. Las que tienen expresión hematimétrica con microcitosis se pueden asemejar a las Talasemias, dificultando a la máquina su diferenciación.

- Muchas de ellas se asocian a una segunda hemoglobinopatía y recordemos que estos pacientes quedaban fuera por los criterios de exclusión.

Estos 3 puntos hacen que obtener hemoglobinopatías estructurales aisladas, que fueran analíticamente visibles, haya sido una tarea muy complicada, clasificándose muchas de estas hemoglobinopatías como pacientes sanos o incluso como paciente con Talasemias (Figura 35).

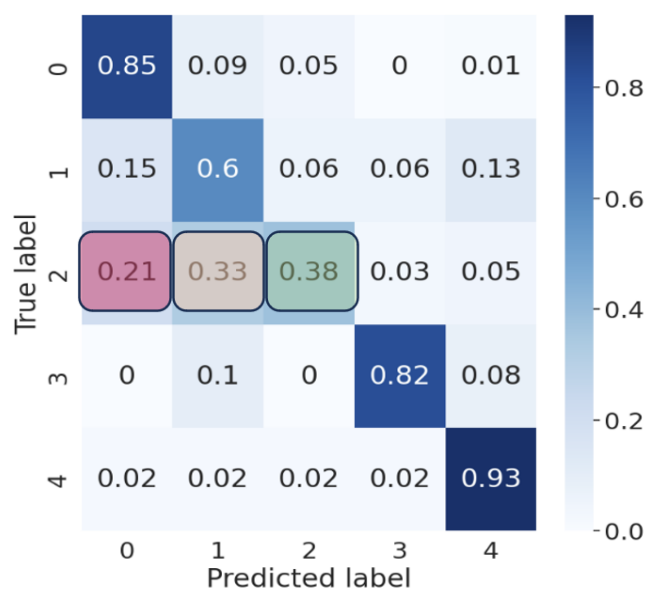


FIGURA 35. MATRIZ DE CONFUSIÓN DE LA RED NEURONAL (ESCENARIO 3). DE COLOR VERDE LAS HEMOGLOBINOPATÍAS ESTRUCTURALES BIEN CLASIFICADAS (38%). DE COLOR NARANJA, LAS HEMOGLOBINOPATÍAS ESTRUCTURALES CLASIFICADAS COMO TALASEMIAS (33%). DE COLOR ROJO LAS CLASIFICADAS COMO PACIENTES SANOS (21%).

CLASE 0 PACIENTES SANOS; CLASE 1: TALASEMIAS, CLASE 2: HEMOGLOBINOPATÍAS ESTRUCTURALES; CLASE 3: ANEMIAS FERROPÉNICAS, CLASE 4: ANEMIAS POR TRASTORNO CRÓNICO.

El problema derivado de eliminar las Hemoglobinopatías Estructurales es que estamos sesgando de alguna manera al sistema, como comentábamos previamente con la publicación de *Obstfeld et al.* (112), ya que estamos eliminando una posible causa de anemia y, por tanto, el sistema no aprenderá a identificarlas. De la misma manera nuestro algoritmo será incapaz de diagnosticar un paciente que se haya transfundido

previamente, por estar ausente su ejemplo entre los aprendidos. En resumen, podemos afirmar que para el diseño de un estudio con participación de IA en salud debemos tener en cuenta la dificultad del problema que nos planteemos para poder decidir la cantidad y la calidad de los datos que necesitamos aportar.

En este sentido una de las dificultades que nos encontramos es la ausencia de directrices que se puedan seguir cuando se plantean proyectos de este tipo. Por ello, para determinar el tamaño muestral y la limpieza de datos es recomendable apoyarse en los resultados obtenidos por otros equipos de investigación en el mismo campo.

Asimismo, además de definir claramente cuáles eran los criterios de exclusión de nuestras muestras, desde el equipo de IA se aplicaron filtros de eliminación de aquellos valores que superaran las 6 desviaciones típicas con respecto a la media de las muestras clasificadas en un mismo grupo de etiquetas (diagnósticos). El valor de estos filtrados ya ha sido publicado por otros grupos de trabajos mostrando mejora en la clasificación de anemias (209).

## 7.3 El valor de aportar conocimiento específico a una Red

### Neuronal

Con intención de mejorar los resultados publicados hemos logrado demostrar que aportar conocimiento específico a una Red Neuronal (BloodNet) ha mejorado los datos de la misma, comparándolos con TabNet, otra Red Neuronal de características y complejidad similares (**Tabla 14**), siendo esta nuestra mayor aportación a la ciencia, ha sido publicada reciente en una revista de primer cuartil del JCR en medicina (202).

Pero, ¿qué significa aportar conocimiento específico a una Red Neuronal? Pues con él, tratamos de explicarle a la máquina cómo se hace una lectura mental del hemograma por un especialista. Inicialmente el hematólogo suele fijarse en la serie roja, la serie

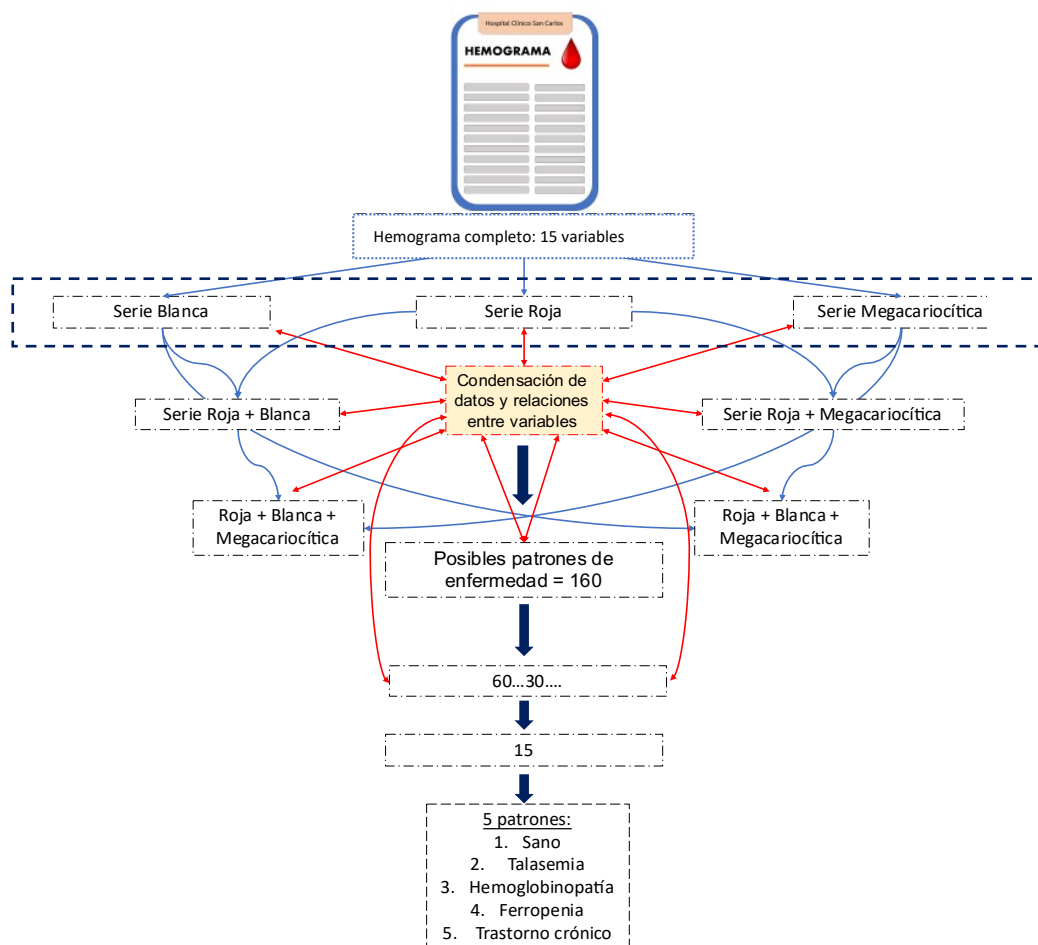


FIGURA 36. REPRESENTACIÓN ILUSTRATIVA DEL PROCESAMIENTO DE LOS DATOS POR BLOODNET.

blanca y la serie megacariocítica de manera independiente y posteriormente hace un repaso de las 3 series y entrelaza sus resultados para valorar un diagnóstico. Esto lo hacemos de manera inconsciente y en milésimas de segundo. Aportarle esta forma de entender el hemograma a un sistema inteligente es la primera vez que se publica. En la **Figura 36** vemos como inicialmente (líneas discontinuas) la red separa las diferentes series hematimétricas y las va entrelazando y comparando entre ellas. Después consigue características de cada hemograma que ordena en patrones de diagnóstico que va condensando hasta llegar a la clasificación entre 5 posibles opciones (pacientes sanos, anemias ferropénicas, anemias por trastorno crónico, talasemias y hemoglobinopatías estructurales).

Aportarle esta manera de razonamiento a BloodNet le ha conferido una mayor capacidad de aprendizaje frente al resto de sistemas de Inteligencia Artificial comparados en esta tesis doctoral. Aunque no es la primera vez que se le aporta conocimiento específico a un algoritmo inteligente, sí es la primera vez que se utiliza para la lectura y comprensión de un hemograma. Como ya hemos comentado en Materiales y Métodos, nuestra idea está inspirada en otra Red Neuronal (COxPASNeT®) en la cual se fusionaban datos clínicos y datos analíticos, pero previamente se realizaba un aprendizaje de ambos grupos de manera independiente (201).

## 7.4 La importancia de la explicabilidad en ciencias de la salud. La caja negra (*The Blackbox*)

Desde el inicio de la aparición de la IA existe escepticismo a la hora de confiar en los diagnósticos que ofrecen las nuevas tecnologías de Inteligencia Artificial y en cierta manera es una opinión fundada, pues ya hemos visto proyectos de IA con errores no permitidos cuando hablamos de salud, como los provocados por el sistema de IA en oncología, *Watson for Oncology* de IBM®, que provocó su retirada de la mayoría de los países donde se utilizaba en la práctica clínica (187).

Uno de los grandes problemas a los que se enfrenta la IA es el desconocimiento de cómo se procesan los datos aportados y este desconocimiento aumenta con las Redes Neuronales de alta complejidad (Deep Learning). Todo esto nos hace plantearnos en sanidad dudas tales como: ¿debemos fiarnos de sus diagnósticos? ¿en base a qué? Y es que la explicabilidad en medicina es obligatoria. Este tema es denominado como la caja negra de la IA (*Blackbox*) y conseguir sistemas de explicabilidad es objetivo de muchas publicaciones (192,194)

En esta tesis doctoral, entender los resultados de nuestra propia Red Neuronal era uno de los objetivos que nos propusimos. Para ello utilizamos sistemas de explicabilidad basados en valores SHAP y gráficos de Beeswarm. El funcionamiento de estos sistemas está explicado en el [apartado 6.5](#).

Los gráficos de explicabilidad nos muestran detalles que se correlacionan con la práctica clínica y los valores de diagnóstico de las anemias. Por ejemplo, vemos que la variable Edad la gana peso y se coloca en el cuarto puesto en el diagnóstico de Anemias de Trastorno Crónico o que los Hematíes (HEMA) y la Hemoglobina Corpuscular Media

(HCM) marcan un gran peso en el diagnóstico de las Talasemias. La importancia de estas variables coincide con lo publicado en la literatura científica (41).

De la misma manera, podemos ver en el *Blot Pot* del escenario 2 (**Figura 27**) que una de las variables de menor peso para el diagnóstico de las anemias han sido los leucocitos totales (WBC). Estos resultados, así como el considerar al HCM como la variable de mayor importancia, son compartidos con los datos publicados por el grupo de *Mo, Donghua et al.* (162).

En definitiva, tener un sistema de explicabilidad que justifique por qué el algoritmo ha decidido un diagnóstico concreto aporta confianza en el sistema inteligente. Desde nuestro punto de vista, una forma de uso real de estos sistemas se representa en la **Figura 37** y su utilidad se puede comparar al de la validación externa de un sistema, ya que, de alguna forma, el médico que recibe el resultado puede comprobar que las variables en las que se ha basado el algoritmo inteligente corresponden a las validadas por el estado del arte en anemias.

## Alta sospecha de Talasemia

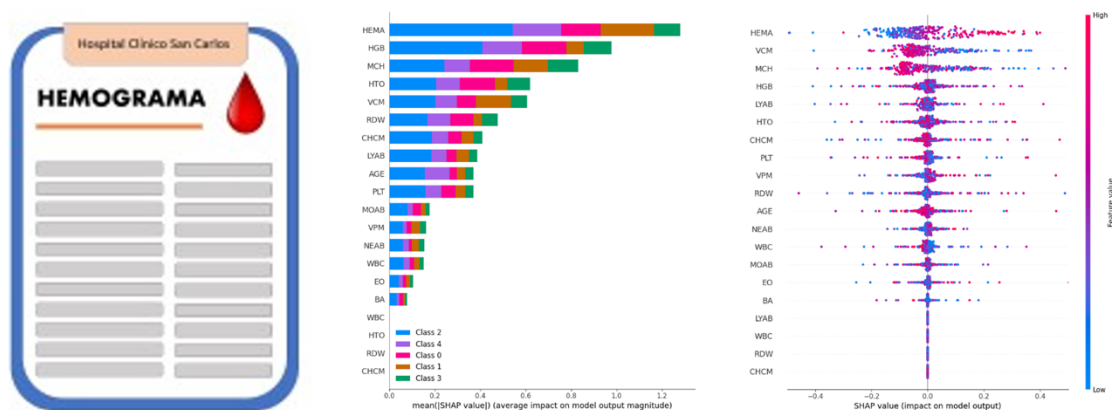


FIGURA 37. REPRESENTACIÓN DE USO REAL DE BLOODNET EJEMPLIFICADO CON UN DIAGNÓSTICO DE TALASEMIA.

## 7.5 La importancia de la validación externa

Otro de los grandes problemas de la IA en general y en ciencias de la salud en concreto es que los resultados obtenidos no se comparan con otras poblaciones, otros hospitales o con diferentes analizadores. Es decir, no se realizan validaciones externas de los algoritmos de IA, lo que se traduce en que ese sistema informático funciona para esa población, en ese laboratorio, con dichos analizadores, sin poder extrapolar sus resultados a otros lugares. Las publicaciones donde no se realizan validaciones externas son frecuentes en medicina y hematología (8,10,161), incluso en aquellas publicaciones posicionadas entre las mejores revistas, la validación externa está ausente (117). Este problema viene condicionado por la dificultad para compartir datos entre hospitales, países y personas y las dificultades burocráticas que se generan alrededor del uso de datos sanitarios de pacientes.

En el diseño de Redes Neuronales siempre se mantiene un pool de muestras sin utilizar en el entrenamiento del sistema. En nuestro caso un 10% de las muestras fueron destinadas a ello, pero además realizamos una segunda recogida de datos desde enero del 2019 a marzo del 2021 para realizar una segunda validación, simulando una puesta a punto en vida real, con lo que reafirmamos que nuestra maquina había aprendido realmente a reconocer las características intrínsecas de cada enfermedad. Para ello, en una segunda muestra de 250 hemogramas cuya distribución de muestra en la **Tabla 8** hemos demostrado que BloodNet funciona en pacientes que no habían sido incluidos para el entrenamiento de la misma (**Figura 33** y **Tabla 13**). No obstante, queda pendiente llevar este sistema de IA a otros hospitales con diferente población para realizar una validación externa estricta.

## 7.6 La Importancia de las variables incluidas en el estudio

Uno de los primeros pasos a la hora de preparar nuestros datos para ser suministrados a un algoritmo predictivo, es el de deshacerse de aquellas variables fuertemente correlacionadas ya que aumentan la dimensionalidad del problema sin aportar gran información. En nuestro primer año de trabajo, a partir de una muestra inicial de 300 pacientes, hicimos una comprobación de relación entre las diferentes variables del hemograma. Los resultados se muestran en la matriz de correlación de la **Figura 38**.

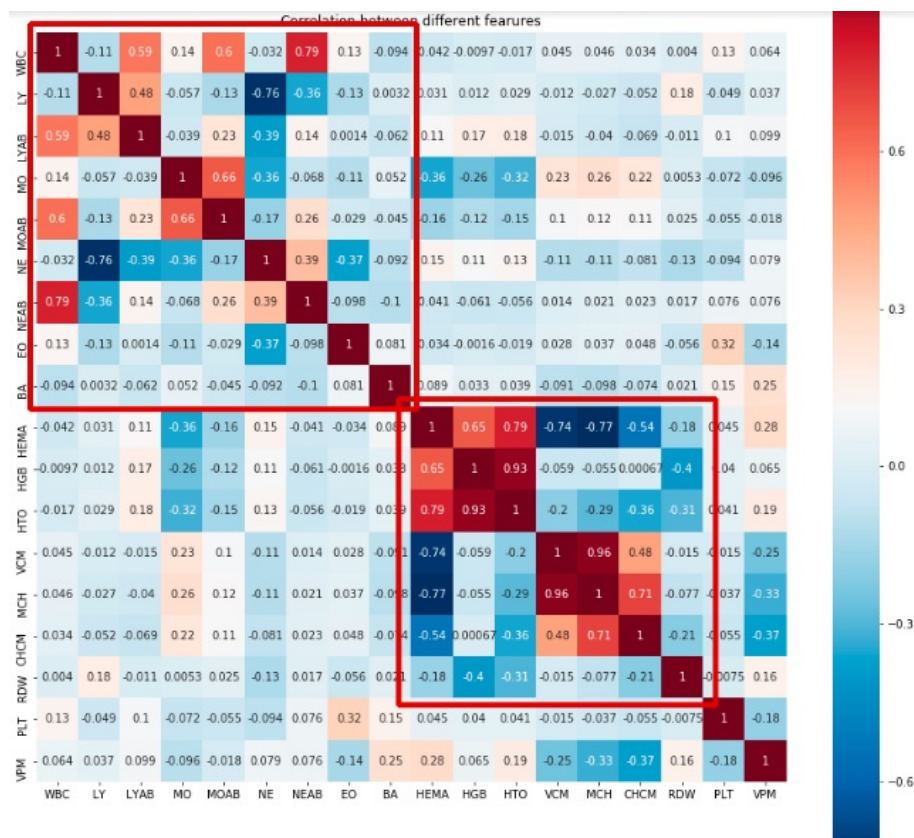


FIGURA 38. MATRIZ DE CORRELACIÓN DE LAS VARIABLES INCLUIDAS EN EL ESTUDIO DE ESTA TESIS DOCTORAL

Como podemos ver en el recuadro inferior-derecho de nuestra figura, entre nuestros hemogramas ocurría lo que vemos a diario en el hospital: existe una fuerte correlación entre Hematocritos (HTO) y la Hemoglobina (HGB).

Estos valores suelen estar fuertemente correlacionados en pacientes sanos, pero es justamente en los pacientes con anemia y alteraciones en su Volumen Corpuscular Medio (VCM) en los que aparecen discrepancias entre ambos valores. Algo parecido ocurre entre el VCM y la Hemoglobina Corpuscular Media (MCH) que, de manera general, van a tener una fuerte correlación. Sin embargo, en un hospital de diagnóstico centralizado (receptor de muestras externas) el VCM puede verse alterado por la conservación y transporte de la muestra, siendo el MCH menos sensible a estas adversidades (210).

Además de esto, queríamos valorar si nuestra Red Neuronal era capaz de encontrar otras variables del hemograma que fueran de interés a la hora de subclassificar las diferentes anemias, pero como mostramos en los gráficos de explicabilidad las variables con mayor peso son de mayor a menor: HCM > HGB > HTO > HEMA > VCM para el escenario 2 (**Figura 27**) y HEMA > HGB > HCM > HTO > VCM en el escenario 3 (**Figura 31**). Esta idea estaba inspirada en otras publicaciones de IA en el diagnóstico de Síndromes Mielodisplásico, donde el algoritmo había sido capaz de encontrar características morfológicas en la médula ósea que orientaban hacia alteraciones genéticas concretas, o estudios donde la imagen de un frotis de sangre periférica de una membranopatía trataba de averiguar la alteración genética causante de su patología, algo prácticamente impensable para el ojo humano (148,177).

Hemos eliminado los resultados en porcentaje de linfocitos, neutrófilos y monocitos para utilizar sus valores absolutos.

## 7.7 Comparando nuestros resultados con lo publicado

### previamente

En 2018 fue publicado por el Centro Médico Universitario de Liubliana (UMCL) la primera publicación con un número elevado de muestras que utilizaban IA en el diagnóstico hematológico, recogiendo 371.341 análisis de 8.233 pacientes con diagnósticos hematológicos entre 2005 y 2015. Entre todos ellos, se incluyeron 1.865 pacientes con anemia (1190 con anemia ferropénica, 457 pacientes con Síndrome Mielodisplásico y 218 con otros tipos de anemia), utilizando métodos tradicionales de Machine Learning para su análisis posterior (117).

Nuestro trabajo se inició ese mismo año 2018 y se recopilaron inicialmente un total de 111.144 hemogramas entre 2011 y 2018, de los cuales etiquetamos 4.060 hemogramas con diagnóstico (incluidos los pacientes sanos) y que tras la limpieza de datos se obtuvieron los 3.295 hemogramas: 863 controles sanos, 675 talasemias, 798 hemoglobinopatías estructurales, 455 anemias ferropénicas y 504 anemias por trastorno crónico (**Tabla 8**). Con estos datos demostramos que la muestra de nuestro trabajo es suficiente para realizar un estudio basado en el análisis de datos con redes neuronales.

Los resultados de dicho trabajo publicado por *Gunčar et al.* mostraron una *exactitud* global de su trabajo del 59% y 56% cuando utilizaron 181 y 65 parámetros analíticos respectivamente. Sin embargo, cuando su algoritmo disminuía la complejidad, y clasificaba las muestras entre las 5 enfermedades hematológicas más frecuentes, la exactitud aumentaba hasta el 90 % a partir de 181 parámetros incluidos y conseguían un 85% cuando incluían 65 parámetros analíticos (117).

En esta tesis doctoral únicamente hemos incluido hemogramas, no otros datos analíticos bioquímicos y por tanto las variables incluidas son reducidas (15 analíticas, además de la Edad y el Sexo). Nuestros resultados medidos en Exactitud/Accuracy son del 71,8% cuando diferenciamos entre 5 enfermedades y del 88,6 % cuando excluimos las Hemoglobinopatías Estructurales. Estos resultados muestran que el aumento del número de variables estudiados no es proporcional a la mejora de los resultados que consiga el algoritmo. Es decir, el aumento de información no siempre se traduce en mejores resultados.

En el trabajo de *Gunčar et al.* se incluyen hasta 181 variables, un 90% más que en esta tesis, aumentando la cantidad de datos a procesar y, por tanto, la complejidad del algoritmo para la clasificación de enfermedades, lo que puede ser contraproducente.

Existen trabajos como el de *Kabootarizadeh Leila et al.* que apoyan este argumento, donde utilizan únicamente 4 variables (HBG, Eritrocitos, VCM y HCM) en un grupo pequeño de 70 pacientes, con un Accuracy del 90% a la hora de diferenciar entre Anemias Ferropénicas y Talasemias, es decir, llevando a cabo una clasificación binaria (163).

A este respecto, debemos aprender a ser críticos con la dificultad de la clasificación que está llevando a cabo un sistema inteligente. No se pueden comparar trabajos de complejidades diferentes y debemos conocer sobre cuantas enfermedades está trabajando el algoritmo inteligente.

La mayoría de los trabajos encontrados en la bibliografía comparan dos únicos grupos, lo que facilita al algoritmo encontrar el diagnóstico con mayor facilidad, llegando a conseguir en todos ellos una precisión mayor al 90% en las clasificaciones (159,162,198).

En esta tesis doctoral nos marcamos el objetivo de aportar al problema una complejidad de vida real al tratar de diagnosticar las cuatro causas de anemia más frecuentes en nuestro medio (además de los pacientes sanos), exigiéndole a BloodNet la dificultad discriminativa propuesta en el *escenario 3* (incluso excluyéndose las HE). Por ello hemos incluido todo tipo de talasemias: alfa, beta, delta-beta y épsilon-gamma-delta-beta talasemias (**Tabla 5**), además de pacientes con Anemia Ferropénica, Anemias por Trastorno Crónico y Pacientes Sanos, lo que hace que nuestro sistema pueda ser útil en vida real.

No hemos encontrado un estudio publicado hasta la fecha similar en cuanto a la variabilidad de anemias incluidas en el estudio con un número de muestras tan elevado, siendo lo más frecuente la publicación de aplicaciones de IA en el screening de Talasemias, diferenciado únicamente entre sanos-enfermos, es decir una clasificación binaria como la de nuestro escenario 1 (159,160,174,211,212,163,164,167–170,172,173).

La excepción a la regla la encontramos en una publicación reciente del grupo de *Saputra et al.* en la que incluyen, además de beta-talasemias y pacientes con anemia ferropénica, pacientes con Hemoglobina E, consiguiendo resultados con una exactitud del 99%, con un número de muestras escaso (190 pacientes, de los cuales son 165 mujeres y 25 hombres) y donde no se ha realizado una validación externa (10).

Para finalizar este punto, cabe recordar que desde la década de 1970 existen publicaciones buscando la mejor fórmula matemática que discrimine Talasemias, pero ninguna de esas fórmulas ha sido introducida en ningún analizador, ya que la mayoría de ellas está basada en un grupo limitado de pacientes (menos de 2500 muestras),

excepto la publicación de 1977 de *Shine et al.* en la que se incluyeron 25.000 muestras, aunque únicamente 138 de esas muestras correspondían a pacientes con Talasemia (213). Esta publicación muestra sensibilidades de hasta el 99,56%, pero con una especificidad muy baja del 22,04%. En este sentido, los resultados de las fórmulas matemáticas han sido superadas en todos los campos por la IA (160). Si vemos los resultados de nuestro trabajo cuando diferenciamos entre pacientes sanos y enfermos en el escenario 1 conseguimos una exactitud del 96%, un Recall (sensibilidad) del 96% y una AUC del 0,96 en las curvas ROC (**Tabla 9, Figura 19 y Figura 20**).

## 7.8 Evaluación coste-efectiva

Los costes para la realización de un algoritmo inteligente están basados en la contratación de personal profesional (ingenieros informáticos y médicos) durante el proceso de la recogida de datos, diseño y entrenamiento del algoritmo inteligente. Posteriormente el coste de cada estudio será únicamente el de la analítica necesaria que mostraremos al sistema: un hemograma, cuyo valor es menor al de 50€. Este coste es muy bajo en comparación con los costes de los estudios genéticos que en ocasiones realizamos por solicitud, cuyo coste puede rondar los 2.000€ por paciente (32,33).

Como comentamos en el [apartado 5.5](#) y mostramos en la **Tabla 6**, en la práctica habitual más de un 30% de las muestras que llegan al centro de referencia para estudio acaban siendo negativas para Hemoglobinopatías, realizándose el estudio genético completo para poder descartarla.

Además, el tiempo de diagnóstico de nuestro actual sistema inteligente está por debajo del minuto (instantáneo) mientras que el estudio habitual de una Hemoglobinopatía puede comprender entre 2 a 5 semanas, según el hospital donde trabajemos (214). A esto tenemos que añadir que el paciente suele acudir derivado desde su Médico de Atención Primaria, lo que alarga más aún los tiempos de diagnóstico.

Por tanto, los resultados de esta tesis representan una opción de mejora al sistema de diagnóstico actual en cuanto a la velocidad del mismo, una posible herramienta de apoyo en el diagnóstico diferencial de anemias orientada a aquellos centros sanitarios donde no se dispone de especialistas y una posible mejora en el uso eficiente de los recursos, tanto materiales como humanos, de los que disponemos.

## 7.9 Limitaciones de esta tesis doctoral

Nuestros resultados han estado limitados por la cantidad de datos obtenidos, de una manera u otra. Podemos resumir nuestras mayores dificultades en cuanto a la recogida de datos en los siguientes puntos:

### 1. Déficit de datos:

- En la mayoría de nuestras muestras no hemos tenido acceso a información sobre otras patologías concomitantes del paciente que pueden alterar el hemograma, por ejemplo, una trombopenia inmune. Es por ello que en los criterios de exclusión fuimos estrictos con estas características.
- No hemos tenido acceso a la raza, algo importante en el diagnóstico de Hemoglobinopatías. Aunque esta tesis se ha llevado a cabo en España, existe una diversidad racial elevada en la comunidad de Madrid y, por tanto, muchos de nuestros pacientes no son caucásicos, siendo hasta de un 15% la población extranjera empadronada en Madrid en los datos publicados en 2021 (215). Esto provoca que nuestro algoritmo sea capaz de averiguar el diagnóstico sin aportarle la raza de cada persona y que, por tanto, tenga una mayor validez externa.
- La elevada variabilidad en el hemograma que presenta el gran abanico de las Hemoglobinopatías Estructurales, siendo nuestro mayor problema de clasificación. Para superar este problema tendríamos que limitar el estudio a un grupo concreto de HE o añadir un número mucho mayor de muestras. Incluso con estas medidas, dado que las HE en ocasiones no tienen expresión en el hemograma, los resultados podrían seguir siendo de baja calidad.

## 2. Asimetría en el número de diagnósticos por enfermedad:

De igual manera que en la estadística clásica, en Inteligencia Artificial debemos tratar de aportar al sistema el mismo número de pacientes de cada clase. En nuestro caso, las enfermedades limitantes por su número en comparación con el resto de enfermedades han sido las Talasemias que no cumplían los criterios de exclusión, teniendo que adaptar la cantidad del resto de datos al número disponible de Talasemias que obtuvimos tras la limpieza de datos (**Tabla 7**). Pese a que la clasificación ha sido manual y hemos tratado de balancear los grupos, se han utilizado funciones de *Sample Weight* para mejorar el balanceo de la muestra, como se ha comentado en el [apartado 5.9](#).

## 7.10 Reflexión personal

A modo de resumen de estos años de trabajo, reflejamos desde una perspectiva holística nuestra recomendación o pasos a seguir para quienes vayan a iniciar un proyecto de Inteligencia Artificial en hematología y/o medicina:

- Identificar una necesidad de mejorar el tratamiento, diagnóstico o screening de una patología a partir de un análisis sencillo y de bajo coste.
- Tener acceso a la base de datos donde recoger la información, a poder ser sin la intervención humana a la hora de recogerlos. Es decir, exportando datos directamente. A tener en cuenta que un 10 % de estos datos serán utilizados únicamente para la validación.
- Estimar la muestra necesaria a partir de los trabajos publicados en Inteligencia Artificial en el mismo campo de estudio.
- Debe existir una simetría en la cantidad de datos que incluimos dentro de las diferentes etiquetas (diagnósticos) y que servirán para el entrenamiento del algoritmo.
- Contactar con un grupo de bioinformática interesado en los datos y/o resultados que quieras conseguir. Recordad que los procesos de anonimización de datos son obligatorios.
- Excluir las muestras artefactadas por tratamientos recibidos, por la conservación de la muestra o por procesos concomitantes que alteren los resultados analíticos al diagnóstico.
- La complejidad del problema propuesto dependerá de la cantidad de datos que seamos capaces de obtener. Podemos valernos en este caso de los criterios de

exclusión, pero recordad que, a mayores criterios de exclusión, menor validez externa.

- Investigar sistemas de explicabilidad que aporten confianza al médico o sanitario que sea potencial usuario del sistema en el futuro. Los sistemas de explicabilidad pueden convertirse en una manera de validación personal del sistema inteligente.
- En la medida de lo posible, realizar una validación externa del sistema con una población diferente. Las validaciones externas tendrán mayor validez cuanto más diferente sea la población de validación de la utilizada para el entrenamiento del sistema (otra muestra > otro hospital > otra ciudad > otro país > otro continente).

## 8. Conclusiones

- 1) Hemos formado un equipo de trabajo multidisciplinar entre médicos e ingenieros informáticos y sus instituciones correspondientes: el Hospital Clínico San Carlos y el Departamento de Computación en la Universidad de Málaga. El contacto ha sido telemático de manera trimestral y presencial anualmente. Ha habido flujo cruzado de aprendizaje para entender las bases de conocimiento del campo del grupo contrario.
- 2) Hemos desarrollado un algoritmo de Inteligencia Artificial basado en Redes Neuronales que hemos denominado BloodNet. Dicho algoritmo ha aprendido a diferenciar entre 3 grupos de anemias (Anemias Ferropénicas, Talasemias, Anemias de Trastorno Crónico) y pacientes sanos (4 grupos en total) con una exactitud del 88,6% y cuyas curvas ROC muestran una AUC que van entre el 0,97 y 0,98 dependiendo del diagnóstico referido, lo que corresponde a un resultado excelente.
- 3) La heterogeneidad de las Hemoglobinopatías Estructurales incluidas en esta tesis doctoral, muchas de ellas con expresión ausente en el hemograma, ha impedido que nuestra Red Neuronal consiga resultados con alta precisión en su diagnóstico.
- 4) A pesar de lo comentado en el punto anterior, la complejidad del problema propuesto en esta tesis es superior a la publicada hasta el momento en el diagnóstico diferencial de anemias basado en IA. Más del 90% de las publicaciones encontradas en la literatura únicamente realizan clasificaciones binarias y son escasos los trabajos donde aumentan a 3 grupos de diagnóstico en

anemias. Nuestro trabajo separa 4 diferentes grupos y mantiene una precisión similar a la publicada.

- 5) A mayor número de diagnósticos incluidos, es decir, a mayor complejidad exigida al sistema, menor es la precisión y exactitud del algoritmo. A menos que dicho aumento de complejidad esté acompañado de un aumento del número de muestras de cada clase o diagnóstico.
- 6) Hemos demostrado que la calidad de los datos incluidos en el entrenamiento es uno de los grandes pilares donde trabajar para obtener un buen rendimiento de la Inteligencia Artificial. Esto podemos observarlo en la comparación de resultados entre los escenarios 3 y 3.5 (diferenciándose uno de otro únicamente en la limpieza de datos).
- 7) Hemos demostrado que los sistemas de IA pueden ser entendidos por los médicos. Para ello, se han utilizado sistemas de explicabilidad (basados en valores SHAP y gráficos de Beeswarm). Estos han demostrado que el sistema inteligente valora el hemograma y decide el diagnóstico de una manera similar a la que realizamos en la clínica habitual. La mayor seguridad para un médico es poder comprobar por sí mismo que el diagnóstico ofrecido por el sistema está basado de manera coherente por los valores de sus variables (en este caso del hemograma).
- 8) Los gráficos de explicabilidad han señalado que las variables de mayor peso para el diagnóstico de anemias son la Hemoglobina Corpuscular Media, la Hemoglobina, el Hematocrito, los Hematíes totales y el Volumen Corpuscular Medio; siendo los Leucocitos totales la variable de menor peso. Esto se correlaciona con lo publicado en la bibliografía.

- 9) El sistema de Inteligencia Artificial no ha encontrado nuevas variables de alto peso del hemograma no utilizadas en la práctica clínica para el diagnóstico diferencial de anemias.
- 10) La aplicación de nuestra Red Neuronal (BloodNet) ha sido superior frente a los algoritmos tradicionales de machine Learning evaluados (*Random Forest, Support Vector Machine* y *XGBoost*).
- 11) Nuestros resultados mediante la aplicación de BloodNet han superado a los conseguidos por Redes Neuronales de acceso público (TabNet). Esta mejora se ha conseguido en gran parte mediante la aportación del conocimiento específico de un hematólogo a nuestra Red Neuronal, algo que no se había realizado hasta la fecha.
- 12) Además de aplicar sistemas de explicabilidad, se ha realizado una validación externa o puesta a punto del sistema con una segunda muestra de 250 hemogramas recogidos entre los años 2019 y 2021 en el HCSC para simular una puesta en producción, donde se han confirmado los resultados y con ello la validez de nuestro sistema.
- 13) Aunque los sistemas de aprendizaje requieren un tiempo considerable para “aprender”, una vez que lo han hecho, la respuesta para dar el diagnóstico de un hemograma es inmediata. Esto, junto a que el hemograma es una prueba rápida y barata incluida dentro de cualquier laboratorio básico, plantea una posible mejora en el uso eficiente de los recursos a partir del uso de aplicaciones inteligentes similares a BloodNet en el sistema sanitario.

## 9. Trabajos y ponencias en congresos que apoyan esta tesis doctoral

- Trelles-Martínez, Roberto, Díaz-del-Pino, Sergio, Trelles Oswaldo, González-Fernández, F.A. “Hacia una medicina de precisión hematológica.” LXI Congreso Nacional SEHH y XXXV congreso nacional de la SETH, 2019. Sesión oral.
- Díaz-del-Pino, Sergio, Trelles-Martínez, Roberto, Perez-Wohlfeil, Esteban, Trelles Oswaldo. “Towards the intelligent diagnosis of hematological diseases.” 4th European Conf. on Translational Bioinf: Biomedical Big Data Supporting Precision Medicine. 2019.
- Trelles-Martínez, Roberto, Diaz-del-Pino, Sergio, González-Fernández, F.A. “Aplicando redes neuronales en el diagnóstico de anemias: la importancia de la explicabilidad.” LXV Congreso Nacional SEHH, XXXIX Congreso Nacional SETH y III congreso Iberoamericano de hematología, 2023. Sesión oral.
- Díaz-del-Pino, Sergio, Trelles-Martínez, Roberto, González-Fernández, F.A. & Guil, Nicolás. “Artificial intelligence to assist specialists in the detection of haematological diseases.” Heliyon, Volume 9, Issue 5, 2023, e15940, ISSN 2405-8440. <https://doi.org/10.1016/j.heliyon.2023.e15940>. (Q2, 3.7).

## 10. Bibliografía

1. McLean E, Cogswell M, Egli I, Wojdyla D, de Benoist B. Worldwide prevalence of anaemia, WHO Vitamin and Mineral Nutrition Information System, 1993–2005. *Public Health Nutr.* 2009/04/01. 2009;12(4):444–54.
2. Kassebaum NJ, Jasrasaria R, Naghavi M, Wulf SK, Johns N, Lozano R, et al. A systematic analysis of global anemia burden from 1990 to 2010. *Blood.* 2014 Jan;123(5):615–24.
3. Dukhi N, Sewpaul R, Sekgala MD, Awe OO. Artificial Intelligence Approach for Analyzing Anaemia Prevalence In Children and Adolescents In Brics Countries: A Review. *Curr Res Nutr FOOD Sci.* 2021;9(1):1–10.
4. Arrizabalaga B, Gonzalez FA, Remacha A. *Eritropatologia* [Internet]. 1st ed. Ambos, editor. Barcelona: Ambos Marketing; 2017. 614 p. Available from: <https://www.libroeritropatologia.com/indice.html>
5. Modell B, Darlison M. Global epidemiology of haemoglobin disorders and derived service indicators. *Bull World Health Organ.* 2008 Jun;86(6):480–7.
6. Yang J, Ding X, Li Y, Lin J. Research on Automatic Wharf Unmanned Gate System Based on Artificial Intelligence. In: *Proceedings of the 2019 7th International Conference on Information Technology: IoT and Smart City.* 2019. p. 521–4.
7. Chen D, Lv Z. Artificial intelligence enabled Digital Twins for training autonomous cars. *Internet Things Cyber-Physical Syst.* 2022;2:31–41.

8. Barosi G, Berzuini A, Quaglini S, Stefanelli M. Anemia: An artificial intelligence project. *Haematologica*. 1985;70(4):358–62.
9. Choi RY, Coyner AS, Kalpathy-Cramer J, Chiang MF, Campbell JP. Introduction to Machine Learning, Neural Networks, and Deep Learning. *Transl Vis Sci Technol*. 2020 Feb;9(2):14.
10. Saputra DCE, Sunat K, Ratnaningsih T. A New Artificial Intelligence Approach Using Extreme Learning Machine as the Potentially Effective Model to Predict and Analyze the Diagnosis of Anemia. *Healthc (Basel, Switzerland)*. 2023 Feb;11(5).
11. Makary MA, Daniel M. Medical error-the third leading cause of death in the US. *BMJ*. 2016 May;353:i2139.
12. Monsalve Torra AE, Soriano Payá A (dir. tes. ., Ruiz Fernández D (dir. tes. . *Sistemas de ayuda a la decisión clínica en enfermedades de diagnóstico complejo*. 2017.
13. Schork NJ. Artificial Intelligence and Personalized Medicine. In: VonHoff DD, Han H, editors. *PRECISION MEDICINE IN CANCER THERAPY*. 2019. p. 265–83.
14. Bini SA. Artificial Intelligence, Machine Learning, Deep Learning, and Cognitive Computing: What Do These Terms Mean and How Will They Impact Health Care? *J Arthroplasty*. 2018 Aug;33(8):2358–61.
15. FundéuRAE. «Inteligencia Artificial» es la expresión del 2022 para la FundéuRAE | FundéuRAE [Internet]. 2022 [cited 2023 Jan 27]. Available from: <https://www.fundeu.es/recomendacion/inteligencia-artificial-es-la-expresion-del-2022-para-la-fundeurae/>

16. Loney M, Chernecky C. Anemia. *Oncol Nurs Forum* [Internet]. 2000 Jul [cited 2023 Jan 31];27(6):951–4. Available from: [https://www.who.int/es/health-topics/anaemia#tab=tab\\_1](https://www.who.int/es/health-topics/anaemia#tab=tab_1)
17. Organization WH. Haemoglobin concentrations for the diagnosis of anaemia and assessment of severity. World Health Organization; 2011.
18. Beutler E, Waalen J. The definition of anemia: what is the lower limit of normal of the blood hemoglobin concentration? *Blood*. 2006 Mar;107(5):1747–50.
19. Chaparro CM, Suchdev PS. Anemia epidemiology, pathophysiology, and etiology in low- and middle-income countries. *Ann N Y Acad Sci*. 2019 Aug;1450(1):15–31.
20. Villegas A. [Hemoglobinopathies in Spanish and emigrant population]. Vol. 23, *Anales de medicina interna (Madrid, Spain : 1984)*. Spain; 2006. p. 203–5.
21. Mañú Pereira M, Corrons JLV. Neonatal haemoglobinopathy screening in Spain. *J Clin Pathol*. 2009 Jan;62(1):22–5.
22. Joyanes B, Moro M, Ropero P, Briceño O, Dulín E, Villegas A. [Screening of hemoglobinopathies in a cohort of newborns in Madrid]. *Med Clin (Barc)*. 2006 Mar;126(8):290–2.
23. Gulbis B, Cotton F, Hansen V, Ferster A, Toppet M, Cochaux P, et al. Prevention of hemoglobinopathies in Brussels: a necessity?. *Rev Med Brux*. 2001 Jun;22(3):133–40.
24. Senger JM, Trenkle JA, St John WD. How to design an anemia management protocol. *ANNA J*. 1998 Apr;25(2):235–40.
25. Brown RG. Normocytic and macrocytic anemias. *Postgrad Med*. 1991

- Jun;89(8):125-132,135-136.
26. Ryan K, Bain BJ, Worthington D, James J, Plews D, Mason A, et al. Significant haemoglobinopathies: guidelines for screening and diagnosis. *Br J Haematol*. 2010 Apr;149(1):35–49.
  27. Grupo Español de Citología Hematológica (GECG). Atlas del GECH [Internet]. 2023 [cited 2023 Jul 18]. Available from: <https://atlas.gechem.org/index.php?lang=es#>
  28. K T N, Prasad K, Singh BMK. Analysis of red blood cells from peripheral blood smear images for anemia detection: a methodological review. *Med Biol Eng Comput*. 2022 Sep;60(9):2445–62.
  29. Cela E, Beléndez C, Galarón P. Interpretación de la electroforesis de hemoglobina. *An Pediatría Contin* [Internet]. 2009;7(3):152–5. Available from: <https://www.elsevier.es/es-revista-anales-pediatria-continuada-51-articulo-interpretacion-electroforesis-hemoglobina-S1696281809711199>
  30. Fucharoen S, Weatherall DJ. The hemoglobin E thalasseмии. *Cold Spring Harb Perspect Med*. 2012 Aug;2(8).
  31. Harteveld CL, Higgs DR. Alpha-thalassaemia. *Orphanet J Rare Dis*. 2010 May;5:13.
  32. Leung KY, Lee CP, Tang MHY, Lau ET, Ng LKL, Lee YP, et al. Cost-effectiveness of prenatal screening for thalassaemia in Hong Kong. *Prenat Diagn*. 2004 Nov;24(11):899–907.
  33. Caballero Villarraso J. Cromatografía líquida desnaturizante de alto rendimiento (DHPLC). Sevilla : Agencia de Evaluación de Tecnologías Sanitarias de Andalucía.; (Ministerio de Sanidad y Consumo. Informes, estudios e Investigación).

34. Weiss G, Ganz T, Goodnough LT. Anemia of inflammation. *Blood*. 2019 Jan 3;133(1):40–50.
35. Berentsen S, Barcellini W. Autoimmune Hemolytic Anemias. *N Engl J Med*. 2021 Oct;385(15):1407–19.
36. Scully M, Hunt BJ, Benjamin S, Liesner R, Rose P, Peyvandi F, et al. Guidelines on the diagnosis and management of thrombotic thrombocytopenic purpura and other thrombotic microangiopathies. *Br J Haematol*. 2012 Aug;158(3):323–35.
37. Brodsky RA. How I treat paroxysmal nocturnal hemoglobinuria. *Blood*. 2021 Mar;137(10):1304–9.
38. Narla J, Mohandas N. Red cell membrane disorders. *Int J Lab Hematol*. 2017 May;39 Suppl 1:47–52.
39. Wickramasinghe SN, Wood WG. Advances in the understanding of the congenital dyserythropoietic anaemias. *Br J Haematol*. 2005 Nov;131(4):431–46.
40. Canatan D, Vives Corrons JL, Piacentini G, Kara F, Keskinliç B, Tezel B, et al. Immigration and screening programs for hemoglobinopathies in Italy, Spain and Turkey. *Acta Biomed*. 2021 Sep;92(4):e2021410.
41. Kohne E. Hemoglobinopathies: clinical manifestations, diagnosis, and treatment. *Dtsch Arztebl Int*. 2011 Aug;108(31–32):532–40.
42. Torrejón MJ, Ortíz-Cabrera NV, M Nieto J, González Fernández FA, Villegas A, Martínez R, et al. Hb Moncloa: A new variant of haemoglobin that interferes in the quantification of Hb A1c. *Clin Biochem*. 2017 Jun;50(9):521–4.
43. Naik RP, Haywood CJ. Sickle cell trait diagnosis: clinical and social implications.

- Hematol Am Soc Hematol Educ Progr. 2015;2015(1):160–7.
44. Calvo-Villas JM, Zapata Ramos MF, Cuesta Tovar J, de la Iglesia Iñigo S, Ropero Gradilla P, Carreter de Granda E, et al. Prevalence of hemoglobinopathies in pregnant women in the Lanzarote health sanitary area. *An Med Interna*. 2006 May;23(5):206–12.
  45. Turing AM. Computing Machinery and Intelligence. *Mind* [Internet]. 1950 Mar 2;59(236):433–60. Available from: <http://www.jstor.org/stable/2251299>
  46. Samuel A. Some Studies in Machine Learning Using the Game of Checkers. II—Recent Progress. *IBM J Res Dev*. 2000 Feb 1;3:206–26.
  47. Martinez-Heras J. Precision, Recall, F1, Accuracy en clasificación - IArtificial.net [Internet]. 2023 [cited 2023 May 5]. Available from: [https://www.iartificial.net/precision-recall-f1-accuracy-en-clasificacion/#Recall\\_Exhaustividad](https://www.iartificial.net/precision-recall-f1-accuracy-en-clasificacion/#Recall_Exhaustividad)
  48. Abonamah AA, Tariq MU, Shilbayeh S. On the Commoditization of Artificial Intelligence. *Front Psychol*. 2021;12:696346.
  49. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. *Nature*. 2021 Aug;596(7873):583–9.
  50. Travis J. ‘New era in digital biology’: AI reveals structures of nearly all known proteins | Science | AAAS [Internet]. 2022 [cited 2023 Mar 20]. Available from: <https://www.science.org/content/article/new-era-digital-biology-ai-reveals-structures-nearly-all-known-proteins>

51. Schwalbe N, Wahl B. Artificial intelligence and the future of global health. *Lancet*. 2020;395(10236):1579–86.
52. Muhsen IN, Shyr D, Sung AD, Hashmi SK. Machine Learning Applications in the Diagnosis of Benign and Malignant Hematological Diseases. *Clin Hematol Int*. 2021 Mar;3(1):13–20.
53. Greener JG, Kandathil SM, Moffat L, Jones DT. A guide to machine learning for biologists. *Nat Rev Mol Cell Biol*. 2022 Jan;23(1):40–55.
54. Masala GL. Artificial Intelligence Techniques applied to Biomedical Patterns. Ardil C, editor. Vol. 12, PROCEEDINGS OF WORLD ACADEMY OF SCIENCE, ENGINEERING AND TECHNOLOGY, VOL 12. 2006. p. 14-+.
55. Doupe P, Faghmous J, Basu S. Machine Learning for Health Services Researchers. *Value Heal J Int Soc Pharmacoeconomics Outcomes Res*. 2019 Jul;22(7):808–15.
56. Breiman L. Random Forests. *Mach Learn*. 2001;45(1):5–32.
57. De Bruyne S, Speeckaert MM, Van Biesen W, Delanghe JR. Recent evolutions of machine learning applications in clinical laboratory medicine. *Crit Rev Clin Lab Sci*. 2021 Mar;58(2):131–52.
58. Cisco Annual Internet Report - Cisco Annual Internet Report (2018–2023) White Paper - Cisco [Internet]. 2020 [cited 2023 Mar 29]. Available from: <https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html>
59. Timothy B. Lee, Timothy B.Lee, Timothy B. Lee. *Ars Technica*. 2017 [cited 2023 Mar 14]. Waymo makes history testing on public roads with no one at the wheel.

Available from: <https://arstechnica.com/cars/2017/11/fully-driverless-cars-are-here/>

60. Rouhiainen L. Inteligencia Artificial. Madrid: Alienta Editorial. 2018.
61. CC C. Google completa la compra de Fitbit: el fabricante de pulseras y relojes deportivos tiene nuevo hogar [Internet]. 2021 [cited 2023 Mar 29]. Available from: <https://www.xatakandroid.com/mercado/google-completa-compra-fitbit-fabricante-pulseras-relojes-deportivos-tiene-nuevo-hogar>
62. van Dis EAM, Bollen J, Zuidema W, van Rooij R, Bockting CL. ChatGPT: five priorities for research. Vol. 614, Nature. England; 2023. p. 224–6.
63. Teknautas. OpenAI libera ChatGPT-4, la versión más “humana” de su Inteligencia Artificial [Internet]. 2023 [cited 2023 Mar 29]. Available from: [https://www.elconfidencial.com/tecnologia/2023-03-14/chatgpt-4-caracteristicas-anunciado\\_3592851/](https://www.elconfidencial.com/tecnologia/2023-03-14/chatgpt-4-caracteristicas-anunciado_3592851/)
64. Azure OpenAI Service: modelos de lenguaje avanzados | Microsoft Azure [Internet]. [cited 2023 Mar 29]. Available from: <https://azure.microsoft.com/es-es/products/cognitive-services/openai-service>
65. Lee P, Bubeck S, Petro J. Benefits, Limits, and Risks of GPT-4 as an AI Chatbot for Medicine. N Engl J Med. 2023 Mar;388(13):1233–9.
66. Owens B. How Nature readers are using ChatGPT. Vol. 615, Nature. England; 2023. p. 20.
67. Vargas Klapp V. Informática médica y gestión en salud - Medical informatics and health management. Rev Hosp Clin Univ Chile. 11(2):170–85.

68. Beam AL, Drazen JM, Kohane IS, Leong TY, Manrai AK, Rubin EJ. Artificial Intelligence in Medicine. *N Engl J Med*. 2023;388(13):1220–1.
69. Finlayson SG, Subbaswamy A, Singh K, Bowers J, Kupke A, Zittrain J, et al. The Clinician and Dataset Shift in Artificial Intelligence. Vol. 385, *The New England journal of medicine*. United States; 2021. p. 283–6.
70. Haug CJ, Drazen JM. Artificial Intelligence and Machine Learning in Clinical Medicine, 2023. *N Engl J Med*. 2023;388(13):1201–8.
71. Liu X, Cruz Rivera S, Moher D, Calvert MJ, Denniston AK, Chan AW, et al. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. *Nat Med*. 2020;26(9):1364–74.
72. Moons KGM, Altman DG, Reitsma JB, Ioannidis JPA, Macaskill P, Steyerberg EW, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): Explanation and elaboration. *Ann Intern Med*. 2015;162(1):W1–73.
73. Fernández Edreira D, Liñares Blanco J, Castiñeiras Galdo B, Maojo García VM, Pazos A. La Inteligencia Artificial como herramienta para la gestión y explotación de datos, informaciones y conocimientos biomédicos en entornos “Big Data” en la nube. *I+S Rev la Soc Española Informática y Salud*. 2021;(143):30–9.
74. Álvarez Vega M, Quirós Mora LM, Cortes Badilla MV. Inteligencia Artificial y aprendizaje automático en medicina. *Rev Médica Sinerg*. 2020;5(8).
75. Pérez Martín J, Díez Vegas FJ (dir. tes. . Cost-effectiveness analysis with new models of artificial intelligence: Medical applications. 2018.

76. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med*. 2019 Jan;25(1):44–56.
77. Ávila de Tomás JF, Mayer Pujadas MA, Quesada Varela VJ. La Inteligencia Artificial y sus aplicaciones en medicina II: importancia actual y aplicaciones prácticas. *Atención primaria Publicación Of la Soc Española Fam y Comunitaria*. 2021;53(1):81–8.
78. Vives Boix V, Ruiz Fernández D (dir. tes. . Modelos neuronales basados en la metaplasticidad para la ayuda al diagnóstico clínico. 2021.
79. Khan AI, Khan M, Khan R. Artificial Intelligence in Point-of-Care Testing. *Ann Lab Med*. 2023 Sep;43(5):401–7.
80. Lanzagorta Ortega D, Carrillo Pérez DL, Carrillo Esper R. Inteligencia Artificial en medicina: presente y futuro. *Gac Med Mex*. 2022;158(1):55–9.
81. Hamet P, Tremblay J. Artificial intelligence in medicine. *Metabolism*. 2017 Apr;69S:S36–40.
82. Ávila de Tomás JF, Mayer Pujadas MA, Quesada Varela VJ. La Inteligencia Artificial y sus aplicaciones en medicina I: introducción antecedentes a la IA y robótica. *Atención primaria Publicación Of la Soc Española Fam y Comunitaria*. 2020;52(10):778–84.
83. Kumar Y, Koul A, Singla R, Ijaz MF. Artificial intelligence in disease diagnosis: a systematic literature review, synthesizing framework and future research agenda. *J Ambient Intell Humaniz Comput*. 2022 Jan;1–28.
84. Kar K. A Computational Probe into the Behavioral and Neural Markers of Atypical

- Facial Emotion Processing in Autism. *J Neurosci*. 2022 Jun 22;42(25):5115 LP – 5126.
85. Li J, Liu L, Sun J, Pei Y, Yang J, Pan H, et al. Diagnosis and Knowledge Discovery of Turner Syndrome Based on Facial Images Using Machine Learning Methods. *IEEE Access*. 2020;8:214866–81.
86. Huang YH, Wei LH, Chen YS. Detection of the prodromal phase of bipolar disorder from psychological and phonological aspects in social media. *arXiv Prepr arXiv171209183*. 2017;
87. Argalious MY, Farag E. Prediction models using machine learning: The focus remains on prevention. *J Clin Anesth*. 2020;67.
88. Kiranyaz S, Ince T, Gabbouj M. Real-Time Patient-Specific ECG Classification by 1-D Convolutional Neural Networks. *IEEE Trans Biomed Eng*. 2016 Mar;63(3):664–75.
89. Johnson KW, Torres Soto J, Glicksberg BS, Shameer K, Miotto R, Ali M, et al. Artificial Intelligence in Cardiology. *J Am Coll Cardiol [Internet]*. 2018;71(23):2668–79. Available from: <https://www.sciencedirect.com/science/article/pii/S0735109718344085>
90. Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, et al. Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. *JAMA*. 2016 Dec 13;316(22):2402–10.
91. Korot E, Pontikos N, Liu X, Wagner SK, Faes L, Huemer J, et al. Predicting sex from

- retinal fundus photographs using automated deep learning. *Sci Rep*. 2021 May;11(1):10286.
92. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*. 2017;542(7639):115–8.
93. Nanaa A, Akkus Z, Lee WY, Pantanowitz L, Salama ME. Machine learning and augmented human intelligence use in histomorphology for haematolymphoid disorders. *Pathology*. 2021 Apr;53(3):400–7.
94. Försch S, Klauschen F, Hufnagl P, Roth W. Artificial Intelligence in Pathology. *Dtsch Arztebl Int*. 2021 Mar;118(12):194–204.
95. Jiang Y, Yang M, Wang S, Li X, Sun Y. Emerging role of deep learning-based artificial intelligence in tumor pathology. *Cancer Commun (London, England)*. 2020 Apr;40(4):154–66.
96. Belén Tobalina. Inteligencia Artificial frente a radiólogos: las máquinas cobran ventaja [Internet]. 2020 [cited 2023 Apr 8]. Available from: <https://www.larazon.es/salud/20201130/nixqxdckrfg3pjoxry5la7nlm.html>
97. McKinney SM, Sieniek M, Godbole V, Godwin J, Antropova N, Ashrafian H, et al. Addendum: International evaluation of an AI system for breast cancer screening. *Nature*. 2020 Oct;586(7829):E19.
98. Moezzi M, Shirbandi K, Shahvandi HK, Arjmand B, Rahim F. The diagnostic accuracy of Artificial Intelligence-Assisted CT imaging in COVID-19 disease: A systematic review and meta-analysis. *Informatics Med unlocked*.

2021;24:100591.

99. Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts HJWL. Artificial intelligence in radiology. *Nat Rev Cancer*. 2018 Aug;18(8):500–10.
100. Zini G. Artificial intelligence in hematology. *HEMATOLOGY*. 2005 Oct;10(5):393–400.
101. Alsuliman T, Humaidan D, Sliman L. Machine learning and artificial intelligence in the service of medicine: Necessity or potentiality? *Curr Res Transl Med*. 2020 Nov;68(4):245–51.
102. Arai Y, Kondo T, Fuse K, Shibasaki Y, Masuko M, Sugita J, et al. Using a machine learning algorithm to predict acute graft-versus-host disease following allogeneic transplantation. *Blood Adv*. 2019 Nov;3(22):3626–34.
103. Shouval R, Fein JA, Savani B, Mohty M, Nagler A. Machine learning and artificial intelligence in haematology. *Br J Haematol*. 2021 Jan;192(2):239–50.
104. Radakovich N, Nagy M, Nazha A. Artificial Intelligence in Hematology: Current Challenges and Opportunities. *Curr Hematol Malig Rep*. 2020 Jun;15(3):203–10.
105. Manimaran M, Arora A, Lovejoy CA, Gao W, Maruthappu M. Role of artificial intelligence and machine learning in haematology. *J Clin Pathol*. 2022 Sep;75(9):585–7.
106. Walter W, Haferlach C, Nadarajah N, Schmidts I, Kühn C, Kern W, et al. How artificial intelligence might disrupt diagnostics in hematology in the near future. *Oncogene*. 2021 Jun;40(25):4271–80.
107. Mosquera Orgueira A, González Pérez MS, Díaz Arias JÁ, Antelo Rodríguez B,

- Alonso Vence N, Bendaña López Á, et al. Survival prediction and treatment optimization of multiple myeloma patients using machine-learning models based on clinical and gene expression data. *Leukemia*. 2021 Oct;35(10):2924–35.
108. Elsayed B, Elshoeibi AM, Elhadary M, Ferih K, Elsabagh AA, Rahhal A, et al. Applications of Artificial Intelligence in Philadelphia-Negative Myeloproliferative Neoplasms. *Diagnostics (Basel, Switzerland)*. 2023 Mar;13(6).
109. Prodan Žitnik I, Černe D, Mancini I, Simi L, Pazzagli M, Di Resta C, et al. Personalized laboratory medicine: a patient-centered future approach. *Clin Chem Lab Med*. 2018 Nov;56(12):1981–91.
110. Wilkes EE. Using machine learning to predict laboratory test results. *Ann Clin Biochem*. 2016 Nov;53(6):712.
111. van der Vorm LN, Hendriks HA, Smits SM. Performance of the CellaVision DC-1 digital cell imaging analyser for differential counting and morphological classification of blood cells. *J Clin Pathol*. 2023 Mar;76(3):194–201.
112. Obstfeld AE. Hematology and Machine Learning. *J Appl Lab Med*. 2023 Jan;8(1):129–44.
113. Salib C, Bhardwaj S, Hussein S, El Jama S, Petersen B, Scigliano E, et al. Digital AI in Hematology - Integration of the Scpio Labs x100 Scanner with Newly Implemented AI Capabilities into Routine Clinical Workflow. *Blood*. 2021;138 MA- (63rd Annual Meeting and Exposition of the American-Society-of-Hematology (ASH) CL-Atlanta, GA PU-AMER SOC HEMATOLOGY PI-WASHINGTON PA-2021 L ST NW, SUITE 900, WASHINGTON, DC 20036 USA).

114. Angehrn Z, Haldna L, Zandvliet AS, Gil Berglund E, Zeeuw J, Amzal B, et al. Artificial Intelligence and Machine Learning Applied at the Point of Care. *Front Pharmacol.* 2020;11.
115. Bransky A, Larsson A, Aardal E, Ben-Yosef Y, Christenson RH. A Novel Approach to Hematology Testing at the Point of Care. *J Appl Lab Med.* 2021 Mar;6(2):532–42.
116. An R, Man Y, Iram S, Kucukal E, Hasan MN, Huang Y, et al. Point-of-care microchip electrophoresis for integrated anemia and hemoglobin variant testing. *Lab Chip.* 2021 Oct;21(20):3863–75.
117. Gunčar G, Kukar M, Notar M, Brvar M, Černelč P, Notar M NM. An application of machine learning to haematological diagnosis. *Sci Rep.* 2018 Jan;8(1):411.
118. El Alaoui Y, Elomri A, Qaraqe M, Padmanabhan R, Yasin Taha R, El Omri H, et al. A Review of Artificial Intelligence Applications in Hematology Management: Current Practices and Future Prospects. *J Med Internet Res.* 2022 Jul;24(7):e36490.
119. Vályi-Nagy I, Peták I. Development and national rollout of electronic decision support systems using artificial intelligence in the field of onco-hematology. *Magy Onkol.* 2019 Dec;63(4):275–80.
120. Xerri L, Brousse N. Decision tree in hematologic diseases: for a rationalization of histologic, phenotypic and genotypic diagnosis. *Ann Pathol.* 1998 Oct;18(4):343–60.
121. Berger J. [Computers in hematology]. *Vnitr Lek.* 1991 May;37(5):507–13.
122. Straka L, Stránský P, Kmoníček M. A decision-support system for hematology. *Sb Lek.* 1998;99(4):483–92.

123. Walters J. An automated hematology workstation. *Am Clin Lab*. 1997 Apr;16(3):9–10.
124. Sivapalaratnam S. Artificial intelligence and machine learning in haematology. Vol. 185, *British journal of haematology*. England; 2019. p. 207–8.
125. Kaestner L. Artificial intelligence meets hematology. *Transfus Apher Sci Off J World Apher Assoc Off J Eur Soc Haemapheresis*. 2020 Dec;59(6):102986.
126. Amendolia SR, Cossu G, Ganadu ML, Golosio B, Masala GL, Mura GM. A comparative study of K-Nearest Neighbour, Support Vector Machine and Multi-Layer Perceptron for Thalassemia screening. *Chemom Intell Lab Syst*. 2003;69(1–2):13–20.
127. Radakovich N, Nagy M, Nazha A. Machine learning in haematological malignancies. *Lancet Haematol*. 2020 Jul;7(7):e541–50.
128. Syed-Abdul S, Firdani RP, Chung HJ, Uddin M, Hur M, Park JH, et al. Artificial Intelligence based Models for Screening of Hematologic Malignancies using Cell Population Data. *Sci Rep*. 2020 Mar;10(1):4583.
129. Rahim F, Kazemnejad A, Jahangiri M, Malehi AS, Gohari K. Diagnostic performance of classification trees and hematological functions in hematologic disorders: an application of multidimensional scaling and cluster analysis. *BMC Med Inform Decis Mak*. 2021 Nov;21(1):313.
130. Ananthkrishnan B, Shaik A, Akhouri S, Garg P, Gadag V, Kavitha MS. Automated Bone Marrow Cell Classification for Haematological Disease Diagnosis Using Siamese Neural Network. *Diagnostics*. 2023;13(1).

131. Zini G, D'Onofrio G. Neural network in hematopoietic malignancies. *Clin Chim Acta*. 2003 Jul;333(2):195–201.
132. Alomari YM, Sheikh Abdullah SNH, Zaharatul Azma R, Omar K. Automatic detection and quantification of WBCs and RBCs using iterative structured circle detection algorithm. *Comput Math Methods Med*. 2014;2014:979302.
133. Wu YY, Huang TC, Ye RH, Fang WH, Lai SW, Chang PY, et al. A Hematologist-Level Deep Learning Algorithm (BMSNet) for Assessing the Morphologies of Single Nuclear Balls in Bone Marrow Smears: Algorithm Development. *JMIR Med informatics*. 2020 Apr;8(4):e15963.
134. Chen L, Liu Y, Xu H, Ma L, Wang Y, Yu L, et al. Touchable cell biophysics property recognition platforms enable multifunctional blood smart health care. *Microsystems Nanoeng*. 2021;7:103.
135. Gasparin AT, Araujo CIF, Schmitt P, Cardoso MR, Perussolo MC, de Jesus TCS, et al. Hilab system, a new point-of-care hematology analyzer supported by the Internet of Things and Artificial Intelligence. *Sci Rep*. 2022 Jun;12(1):10409.
136. Bachar N, Benbassat D, Brailovsky D, Eshel Y, Gluck D, Levner D, et al. An artificial intelligence-assisted diagnostic platform for rapid near-patient hematology. *Am J Hematol*. 2021;96(10):1264–74.
137. Matek C, Krappe S, Münzenmayer C, Haferlach T, Marr C. Highly accurate differentiation of bone marrow cell morphologies using deep neural networks on a large image data set. *Blood*. 2021 Nov;138(20):1917–27.
138. Kutlu H, Avci E, Özyurt F. White blood cells detection and classification based on

- regional convolutional neural networks. *Med Hypotheses*. 2020 Feb;135:109472.
139. Li N, Fan L, Xu H, Zhang X, Bai Z, Li M, et al. An AI-Aided Diagnostic Framework for Hematologic Neoplasms Based on Morphologic Features and Medical Expertise. *Lab Invest*. 2023 Apr;103(4):100055.
140. Mallesh N, Zhao M, Meintker L, Höllein A, Elsner F, Lüling H, et al. Knowledge transfer to enhance the performance of deep learning models for automated classification of B cell neoplasms. *Patterns (New York, NY)*. 2021 Oct;2(10):100351.
141. Zhao M, Mallesh N, Höllein A, Schabath R, Haferlach C, Haferlach T, et al. Hematologist-Level Classification of Mature B-Cell Neoplasm Using Deep Learning on Multiparameter Flow Cytometry Data. *Cytom Part A J Int Soc Anal Cytol*. 2020 Oct;97(10):1073–80.
142. Duetz C, Bachas C, Westers TM, van de Loosdrecht AA. Computational analysis of flow cytometry data in hematological malignancies: future clinical practice? *Curr Opin Oncol*. 2020 Mar;32(2):162–9.
143. Flores-Montero J, Sanoja-Flores L, Paiva B, Puig N, García-Sánchez O, Böttcher S, et al. Next Generation Flow for highly sensitive and standardized detection of minimal residual disease in multiple myeloma. *Leukemia*. 2017 Oct;31(10):2094–103.
144. Nwanosike EM, Conway BR, Merchant HA, Hasan SS. Potential applications and performance of machine learning techniques and algorithms in clinical practice: A systematic review. *Int J Med Inform*. 2022;159.

145. Ren L, Xu B, Xu J, Li J, Jiang J, Ren Y, et al. A Machine Learning Model to Predict Survival and Therapeutic Responses in Multiple Myeloma. *Int J Mol Sci.* 2023 Apr;24(7).
146. Shouval R, Labopin M, Bondi O, Mishan-Shamay H, Shimoni A, Ciceri F, et al. Prediction of Allogeneic Hematopoietic Stem-Cell Transplantation Mortality 100 Days After Transplantation Using a Machine Learning Algorithm: A European Group for Blood and Marrow Transplantation Acute Leukemia Working Party Retrospective Data Mining Study. *J Clin Oncol Off J Am Soc Clin Oncol.* 2015 Oct;33(28):3144–51.
147. Cazzola M, Sehn LH. Developing a classification of hematologic neoplasms in the era of precision medicine. *Blood.* 2022;140(11):1193–9.
148. Haferlach T. Human and artificial intelligence to illuminate MDS. *Blood.* 2020 Nov;136(20):2243–4.
149. Kim GYE, Noshad M, Stehr H, Rojansky R, Gratzinger D, Oak J, et al. Machine Learning Predictability of Clinical Next Generation Sequencing for Hematologic Malignancies to Guide High-Value Precision Medicine. *AMIA . Annu Symp proceedings AMIA Symp.* 2021;2021:641–50.
150. Sibinga CTS. Artificial intelligence in transfusion medicine and its impact on the quality concept. *Transfus Apher Sci.* 2020;59(6).
151. Al-Hakeim HK, Najm AH, Moustafa SR, Maes M. Construction of an exposure-pathway-phenotype in children with depression due to transfusion-dependent thalassemia: Results of (un)supervised machine learning. *J Affect Disord.* 2021;282:644–55.

152. Islam MM, Rahman MJ, Roy DC, Islam MM, Tawabunnahar M, Ahmed NAMF, et al. Risk Factors Identification and Prediction of Anemia among Women in Bangladesh using Machine Learning Techniques. *Curr Womens Health Rev.* 2022;18(1).
153. Dejene BE, Abuhay TM, Bogale DS. Predicting the level of anemia among Ethiopian pregnant women using homogeneous ensemble machine learning algorithm. *BMC Med Inform Decis Mak.* 2022 Sep;22(1):247.
154. Meena K, Tayal DK, Gupta V, Fatima A. Using classification techniques for statistical analysis of Anemia. *Artif Intell Med.* 2019 Mar;94:138–52.
155. Ohmoto A, Fuji S. Current status of drug repositioning in hematology. *Expert Rev Hematol.* 2021;14(11):1005–11.
156. Passamonti F, Corrao G, Castellani G, Mora B, Maggioni G, Gale RP, et al. The future of research in hematology: Integration of conventional studies with real-world data and artificial intelligence. *Blood Rev.* 2022;54.
157. Rashid MBMA, Chow EKH. Artificial Intelligence-Driven Designer Drug Combinations: From Drug Development to Personalized Medicine. *SLAS Technol.* 2019 Feb;24(1):124–5.
158. Quaglini S, Stefanelli M, Barosi G, Berzuini A. ANEMIA: an expert consultation system. *Comput Biomed Res.* 1986 Feb;19(1):13–27.
159. Zhang F, Yang J, Wang Y, Cai M, Ouyang J, Li J. TT@MHA: A machine learning-based webpage tool for discriminating thalassemia trait from microcytic hypochromic anemia patients. *Clin Chim Acta.* 2023;545:117368.

160. Roth IL, Lachover B, Koren G, Levin C, Zalman L, Koren A, et al. Detection of  $\beta$ -thalassemia carriers by red cell parameters obtained from automatic counters using mathematical formulas. *Mediterr J Hematol Infect Dis.* 2018;10(1):e2018008.
161. Permanente K, Region N, Information B, Information B, Management I, Information B. An Expert System to Diagnose Anemia and Report Results Directly on Hematology Forms. 1996;26(29):16–26.
162. Mo D, Zheng Q, Xiao B, Li L. Predicting thalassemia using deep neural network based on red blood cell indices. *Clin Chim Acta.* 2023 Apr;543:117329.
163. Kabootarizadeh L, Jamshidnezhad A, Koohmareh Z. Differential Diagnosis of Iron-Deficiency Anemia from beta-Thalassemia Trait Using an Intelligent Model in Comparison with Discriminant Indexes. *Acta Inform Med.* 2019 Jun;27(2):78–84.
164. Masala GL, Golosio B, Cutzu R, Pola R. A two-layered classifier based on the radial basis function for the screening of thalassaemia. *Comput Biol Med.* 2013 Nov;43(11):1724–31.
165. Jahangiri M, Rahim F, Saki N, Saki Malehi A. Application of Bayesian Decision Tree in Hematology Research: Differential Diagnosis of  $\beta$ -Thalassemia Trait from Iron Deficiency Anemia. *Comput Math Methods Med.* 2021;2021:6401105.
166. Feng P, Li Y, Liao Z, Yao Z, Lin W, Xie S, et al. An online alpha-thalassemia carrier discrimination model based on random forest and red blood cell parameters for low HbA(2) cases. *Clin Chim Acta.* 2022 Jan;525:1–5.
167. Das R, Saleh S, Nielsen I, Kaviraj A, Sharma P, Dey K, et al. Performance analysis

- of machine learning algorithms and screening formulae for  $\beta$ -thalassemia trait screening of Indian antenatal women. *Int J Med Inform.* 2022 Nov;167:104866.
168. Phirom K, Charoenkwan P, Shoombuatong W, Charoenkwan P, Sirichotiyakul S, Tongsong T. DeepThal: A Deep Learning-Based Framework for the Large-Scale Prediction of the  $\alpha$ (+)-Thalassemia Trait Using Red Blood Cell Parameters. *J Clin Med.* 2022 Oct;11(21).
169. Das R, Datta S, Kaviraj A, Sanyal SN, Nielsen P, Nielsen I, et al. A decision support scheme for beta thalassemia and HbE carrier screening. *J Adv Res.* 2020 Jul;24:183–90.
170. Laengsri V, Shoombuatong W, Adirojananon W, Nantasenamat C, Prachayasittikul V, Nuchnoi P. ThalPred: a web-based prediction tool for discriminating thalassemia trait and iron deficiency anemia. *BMC Med Inform Decis Mak.* 2019 Nov;19(1):212.
171. Yilmaz Z, Bozkurt MR. Determination of women iron deficiency anemia using neural networks. *J Med Syst.* 2012 Oct;36(5):2941–5.
172. Fu YK, Liu HM, Lee LH, Chen YJ, Chien SH, Lin JS, et al. The tvgh-nycu thal-classifier: Development of a machine-learning classifier for differentiating thalassemia and non-thalassemia patients. *Diagnostics.* 2021;11(9).
173. Ambayya A, Sahibon S, Yang TW, Zhang QY, Hassan R, Sathar J. A Novel Algorithm Using Cell Population Data (VCS Parameters) as a Screening Discriminant between Alpha and Beta Thalassemia Traits. *Diagnostics (Basel, Switzerland).* 2021 Nov;11(11).

174. Zhou Y, Zhang J, Wang C, Zhou LL, Zhou LL, Ou D, et al. Application of HbA2 levels and red cell indices-based new model in the differentiation of thalassemia traits from iron deficiency in hypochromic microcytic anemia Cases. *Int J Lab Hematol*. 2020 Oct;42(5):526–32.
175. Lee SY, Chen CME, Lim EYP, Shen L, Sathe A, Singh A, et al. Image Analysis Using Machine Learning for Automated Detection of Hemoglobin H Inclusions in Blood Smears - A Method for Morphologic Detection of Rare Cells. *J Pathol Inform*. 2021;12:18.
176. Salman Khan M, Ullah A, Khan KN, Riaz H, Yousafzai YM, Rahman T, et al. Deep Learning Assisted Automated Assessment of Thalassaemia from Haemoglobin Electrophoresis Images. *Diagnostics (Basel, Switzerland)*. 2022 Oct;12(10).
177. Simionato G, Hinkelmann K, Chachanidze R, Bianchi P, Fermo E, van Wijk R, et al. Red blood cell phenotyping from 3D confocal images using artificial neural networks. *PLoS Comput Biol*. 2021 May;17(5):e1008934.
178. Acharya S, Swaminathan D, Das S, Kansara K, Chakraborty S, Kumar R D, et al. Non-Invasive Estimation of Hemoglobin Using a Multi-Model Stacking Regressor. *IEEE J Biomed Heal informatics*. 2020 Jun;24(6):1717–26.
179. Dimauro G, Griseta ME, Camporeale MG, Clemente F, Guarini A, Maglietta R. An intelligent non-invasive system for automated diagnosis of anemia exploiting a novel dataset. *Artif Intell Med*. 2023;136.
180. Muthalagu R, Bai VT, Gracias D, John S. Developmental screening tool: Accuracy and feasibility of non-invasive anaemia estimation. *Technol Heal Care*. 2018;26(4):723–7.

181. Mitani A, Huang A, Venugopalan S, Corrado GS, Peng L, Webster DR, et al. Detection of anaemia from retinal fundus images via deep learning. *Nat Biomed Eng.* 2020 Jan;4(1):18–27.
182. Mannino RG, Myers DR, Tyburski EA, Caruso C, Boudreaux J, Leong T, et al. Smartphone app for non-invasive detection of anemia using only patient-sourced photos. *Nat Commun.* 2018;9(1).
183. Zhang A, Lou J, Pan Z, Luo J, Zhang X, Zhang H, et al. Prediction of anemia using facial images and deep learning technology in the emergency department. *Front public Heal.* 2022;10:964385.
184. Kwon JM, Cho Y, Jeon KH, Cho S, Kim KH, Baek SD, et al. A deep learning algorithm to detect anaemia with ECGs: a retrospective, multicentre study. *Lancet Digit Heal.* 2020 Jul;2(7):e358–67.
185. Sarsam SM, Al-Samarraie H, Alzahrani AI, Shibghatullah AS. A non-invasive machine learning mechanism for early disease recognition on Twitter: The case of anemia. *Artif Intell Med.* 2022 Dec;134:102428.
186. Kaestner L, Bianchi P. Trends in the Development of Diagnostic Tools for Red Blood Cell-Related Diseases and Anemias. *Front Physiol.* 2020;11:387.
187. Ross C & S. IBM's Watson recommended "unsafe and incorrect" cancer treatments - STAT [Internet]. [cited 2023 Apr 4]. Available from: <https://www.statnews.com/2018/07/25/ibm-watson-recommended-unsafe-incorrect-treatments/>
188. Jie Z, Zhiying Z, Li L. A meta-analysis of Watson for Oncology in clinical application.

- Sci Rep. 2021 Mar;11(1):5792.
189. Liebers N, Richter S, Proynova R, Haux C, Mammen J, Czernilofsky F, et al. A big data repository used for an artificial intelligence approach in outcome prediction of hematological Neoplasms: The BRAIN project. *Oncol Res Treat.* 2021;44(SUPPL 2):95–6.
  190. Cabitza F, Rasoini R, Gensini GF. Unintended Consequences of Machine Learning in Medicine. *JAMA.* 2017 Aug;318(6):517–8.
  191. Watson DS, Krutzinna J, Bruce IN, Mcinnes IB, Barnes MR. Clinical applications of machine learning algorithms : beyond the black box. 2019;886(March):10–3.
  192. Yang G, Ye Q, Xia J. Unbox the black-box for the medical explainable AI via multi-modal and multi-centre data fusion: A mini-review, two showcases and beyond. *Inf Fusion.* 2022 Jan;77:29–52.
  193. Doshi-Velez F, Kim B. Towards A Rigorous Science of Interpretable Machine Learning. 2017;(ML):1–13. Available from: <http://arxiv.org/abs/1702.08608>
  194. Holzinger A, Langs G, Denk H, Zatloukal K, Müller H. Causability and explainability of artificial intelligence in medicine. *Wiley Interdiscip Rev Data Min Knowl Discov.* 2019;9(4):e1312.
  195. Samek W, Wiegand T, Müller KR. Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models. *arXiv Prepr arXiv170808296.* 2017;
  196. Gegúndez Fernández JA. Tecnificación versus humanización: La Inteligencia Artificial al servicio del diagnóstico médico. *Arch Soc Esp Oftalmol.*

- 2018;93(3):1017–9.
197. Agramonte Mayhua I, Chaco Huamani A, Valdiviezo Tovar A, Ramos Challa M. Aplicación de los árboles de decisión en el diagnóstico de Anemia en niños de la ciudad de Arequipa. *Rev Innovación y Softw.* 2022;3(2):26–39.
  198. Ferih K, Elsayed B, Elshoeibi AM, Elsabagh AA, Elhadary M, Soliman A, et al. Applications of Artificial Intelligence in Thalassemia: A Comprehensive Review. *Diagnostics (Basel, Switzerland)*. 2023 Apr;13(9).
  199. Secretaría General del Servicio Madrileño de Salud. Consejería de Sanidad. Memoria año 2018. 2019 [cited 2023 May 22]. p. 6–7 Memorias e Informes del Servicio Madrileño de Salud | Comunidad de Madrid. Available from: <https://www.comunidad.madrid/servicios/salud/memorias-e-informes-servicio-madrileno-salud>
  200. Ropero P, González Fernández FA, Martínez Nieto J, Torres Jiménez WM, Benavente Cuesta C.  $\beta$ -Intermediate Thalassemia: triplication of genes  $\alpha / \beta$  thalassemia heterozygous in Spain. *An Ranm.* 2021;138(138(01)):60–71.
  201. Hao J, Kim Y, Mallavarapu T, Oh JH, Kang M. Cox-PASNet: Pathway-based Sparse Deep Neural Network for Survival Analysis. 2018.
  202. Diaz-Del-Pino S, Trelles-Martinez R, González-Fernández FA, Guil N. Artificial intelligence to assist specialists in the detection of haematological diseases. *Heliyon.* 2023 May;9(5):e15940.
  203. Lundberg SM, Lee SI. A Unified Approach to Interpreting Model Predictions. In: Guyon I, Luxburg U Von, Bengio S, Wallach H, Fergus R, Vishwanathan S, et al.,

- editors. *Advances in Neural Information Processing Systems*. Curran Associates, Inc.; 2017.
204. Martínez Pérez JA, Pérez Martin PS. La curva ROC. *Med Fam Semer* [Internet]. 2023;49(1). Available from: <https://www.elsevier.es/es-revista-medicina-familia-semergen-40-articulo-la-curva-roc-S1138359322001952>
205. Chai SY, Hayat A, Flaherty GT. Integrating artificial intelligence into haematology training and practice: Opportunities, threats and proposed solutions. *Br J Haematol*. 2022 Sep;198(5):807–11.
206. Kaestner L. Artificial intelligence: Training the trainer. *Br J Haematol*. 2022 Sep;198(5):805–6.
207. Walter W, Pohlkamp C, Meggendorfer M, Nadarajah N, Kern W, Haferlach C, et al. Artificial intelligence in hematological diagnostics: Game changer or gadget? *Blood Rev*. 2022 Oct;101019.
208. Tellez D, Litjens G, Bándi P, Bulten W, Bokhorst JM, Ciompi F, et al. Quantifying the effects of data augmentation and stain color normalization in convolutional neural networks for computational pathology. *Med Image Anal*. 2019;58:101544.
209. Terzi E, Saribacak B, Sağlam F, Cengiz MA, Saribacak B, Sağlam F, et al. A Novel Expert System for Diagnosis of Iron Deficiency Anemia. *Comput Math Methods Med*. 2022;2022:7352096.
210. Quintó L, Aponte JJ, Menéndez C, Sacarlal J, Aide P, Espasa M, et al. Relationship between haemoglobin and haematocrit in the definition of anaemia. *Trop Med Int Health*. 2006 Aug;11(8):1295–302.

211. Xu W, Song Y, Zou T, IEEE. Prediction of Thalassemia Based on SAELM Hybrid Algorithm. 2019 3RD INTERNATIONAL CONFERENCE ON DATA SCIENCE AND BUSINESS ANALYTICS (ICDSBA 2019). 2019. p. 227–30.
212. Erten M, Tuncer T. Automated differential diagnosis method for iron deficiency anemia and beta thalassemia trait based on iterative Chi2 feature selector. *Int J Lab Hematol*. 2022 Apr;44(2):430–6.
213. Shine I, Lal S. A strategy to detect beta-thalassaemia minor. *Lancet (London, England)*. 1977 Mar;1(8013):692–4.
214. NHS Foundation Trust. Haemoglobinopathies testing - Oxford University Hospitals [Internet]. 2021 [cited 2023 May 22]. Available from: <https://www.ouh.nhs.uk/services/referrals/genetics/genetics-laboratories/rare-disease-genomics/non-malignant-haematology/haemoglobinopathies.aspx>
215. Comunidad de Madrid. La población extranjera se incrementó un 1,1% el pasado año, con 11.338 residentes más | Comunidad de Madrid [Internet]. 2021. Available from: <https://www.comunidad.madrid/noticias/2021/06/26/poblacion-extranjera-incremento-11-pasado-ano-11338-residentes>