



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA



Escola Tècnica
Superior d'Enginyeria
Informàtica

Escola Tècnica Superior d'Enginyeria Informàtica
Universitat Politècnica de València

Diseño de software para obtener la interacción toxicológica entre fármacos en pacientes con enfermedades crónicas

Trabajo Fin de Grado

Grado en Ingeniería Informática

Autor: Sergio Vázquez Bautista

Tutor: Juan Miguel García Gómez

Directora Experimental: María del Mar Álvarez Torres

2020-2021

Diseño de software para obtener la interacción toxicológica entre fármacos en pacientes con enfermedades crónicas



Agradecimientos

Quiero aprovechar este apartado para agradecer principalmente a Juanmi y la UPV por enseñarme el mundo de la bioinformática. Juanmi me ha dado la oportunidad desde cero de hacer un proyecto que puede tener un impacto en la vida de muchas personas. No sólo él me ha ayudado en este ámbito, sino que Mari, también nos ha enseñado de la mejor forma posible en este campo de la bioinformática en los laboratorios. Gracias a esta asignatura también he encontrado mi vocación dentro de la informática y de nuevo gracias a todos los que han hecho posible que asignaturas como estas se impartan en la UPV.

En otro párrafo, pero no menos importante, tengo que agradecer a mi pareja Irene todo lo que ha hecho por mí, desde discutir ideas con ella, buscar contactos para que nos donen datos y poder llevar mi trabajo a cabo, hasta alimentarme para que no desfalleciera mientras trabajaba a jornada completa, además de entrenar e ir a las competiciones deportivas de élite, y por supuesto, hacer este proyecto, te amo.

A mis padres, como no iban a estar aquí, sin la confianza y los valores que me han enseñado no creo que estuviera escribiendo esto, siempre me habéis apoyado y soy quién soy en gran parte gracias a ellos, os quiero.

A esa persona por la cual en gran parte hago este proyecto, Deni, no tengo más que admiración y agradecimiento eterno, por todo lo que ha luchado y estoy seguro que lo seguirá haciendo. Espero de corazón que este proyecto sirva a muchas personas como ella, que padecen artritis reumatoide y tengan una vida mejor, porque si son la mitad de buena persona que ella, se merecen la mejor calidad de vida posible.

Resumen

Actualmente existe una falta de información en cuanto a la interacción toxicológica entre fármacos y la ausencia de herramientas que puedan ofrecer esta información. El objetivo específico del proyecto es predecir los niveles de toxicidad hepática máximos que no causen un perjuicio en el paciente, valorando factores implicados como el tratamiento, consumo de alcohol, tabaco, sexo y edad, entre otros factores. Proporcionando la máxima calidad de vida y mejorando el plan de actuación médico.

En el presente proyecto se estudia la toxicidad de la interacción entre fármacos proporcionados durante el tratamiento de pacientes con enfermedades crónicas. En particular se analizará el nivel de toxicidad hepático, aumento de colesterol y triglicéridos causado por la interacción entre fármacos. Los resultados derivados de este proyecto serán de utilidad clínica, ya que facilitarán la prevención de dificultades desde el inicio de un tratamiento y durante el mismo, permitiendo una mejora de la calidad de vida del paciente. Además, este proyecto aporta información relevante a los clínicos que ayuda en la mejor elección del tratamiento valorando las diferentes opciones existentes. Todo esto mediante el análisis de Real World Evidence (RWE) y utilizando redes neuronales, lo que facilitará crear herramientas aplicables en el campo de la salud que permitan desarrollar terapias dirigidas, tomar decisiones, hacer un análisis de variables, un análisis de tratamientos, un mejor diagnóstico precoz y optimizar recursos.

Palabras clave: toxicidad, calidad de vida, diagnóstico precoz, hepatotoxicidad, artritis.

Abstract

There is currently a lack of information regarding the toxicological interaction between drugs and the absence of tools which can grant this information. The goal of this project is to provide the maximum quality of life and improve the medical action plan by predicting the maximum levels of liver toxicity that do not harm the patient, whilst assessing the factors involved, such as the treatment itself, alcohol consumption, smoking, sex, age, etc.

This project studies the toxicity of the interaction between drugs provided during the treatment of patients with chronic diseases and it particularly analyses the level of liver toxicity, increased cholesterol and triglycerides caused by drug interaction. The results derived from this project will be clinically useful, as they are meant to help prevent difficulties from the beginning of the treatment and during its process as well, thus allowing for an improvement in the patient's quality of life. Furthermore, this project provides clinical staff with relevant information that is helpful during the selection process of the treatment by assessing the range of options available. In fact, all of this is carried out through the analysis of RWE (Real Word Evidence) and using neural networks, which contribute to the creation of tools that can be applied to the health department that can enable the development of targeted therapies, decision making, analysis of variables and treatments, better early diagnosis and optimization of resources.

Keywords: toxicity, quality of life, early diagnosis, hepatotoxicity, arthritis.

Índice de contenidos

Agradecimientos	3
Resumen	5
Abstract	7
Índice de contenidos	9
Memoria	13
1. Introducción	15
1.1 Motivación	15
1.2 Artritis reumatoide	15
1.3 Prevalencia	16
1.4 Aprendizaje automático	17
1.5 Diagnóstico	18
1.6 Tratamiento	19
1.7 Hipótesis	24
1.8 Objetivos	25
1.9 Resumen de apartados restantes	25
2. Estado del arte	27
2.1 Contexto	27
2.2 Algoritmos de obtención toxicológica	27
2.3 Real World Evidence	28

3. Análisis del problema	30
3.1 Marco legal y ético	30
3.1.1 Justificación científica y ética.....	30
3.1.2 Metodología	31
3.1.3 Información que recibirán los participantes	33
3.1.4 Procedimiento de gestión de los datos personales	33
3.1.5 Compensaciones económicas	34
3.2 Conjunto de datos	34
3.2.1 Descripción de los datos	35
3.3 Equipamiento hardware	39
3.3 Equipamiento software	40
4. Métodos	41
4.1 Arquitecturas de redes neuronales y diseño detallado	41
4.2 Evaluación	49
5. Desarrollo de la solución propuesta y resultados	50
5.1 Predicción toxicológica	50
5.1.1 Resultados con modelo simple	50
5.1.2 Resultados con modelo utilizando capas ocultas	52
5.2 Análisis de resultados	55
6. Conclusiones y reflexiones	59
6.1 Conclusiones	59
6.2 Reflexión sobre el trabajo	59
7. Relación del trabajo con los estudios cursados	61



8. Trabajo futuros	62
9. Bibliografía	63
10. Lista de figuras	68
11. Lista tablas	70



Memoria

1. Introducción

En este capítulo se introducirán los conceptos a tener en cuenta con el tipo de enfermedad crónica, el diagnóstico, y el tratamiento médico que conlleva esta enfermedad, así como los objetivos y la motivación de realizar este trabajo.

1.1 Motivación

Este proyecto nace como muchos otros de una necesidad, una necesidad sobre la cual a día de hoy no hay herramientas útiles. La toxicidad que puede desarrollar un paciente, combinando diferentes tipos de fármacos se desconoce completamente cuando hablamos de enfermedades crónicas, ya que utilizan una gran variedad de medicamentos, teniendo que probar con un tipo de medicación u otra sin tener la certeza de que tomar esta gran cantidad de fármacos no conlleva un mal mayor que padecer la enfermedad.

Por lo comentado anteriormente y sumado a que un familiar cercano padece una enfermedad crónica y ha de medicarse diariamente con 11 tipos de fármacos diferentes casi a diario, hemos decidido hacer el proyecto basándonos en la artritis reumatoide. Hemos elegido una enfermedad que nos proporciona unos valores hepáticos claros, que son los que vamos a tener en cuenta para calcular/prever la interacción toxicológica. Estos valores podemos tenerlos en cuenta ya que actualmente se desconoce si la artritis reumatoide tiene afección hepática directa.

1.2 Artritis reumatoide

La artritis reumatoide es una enfermedad inflamatoria, autoinmune y sistémica que se caracteriza por la inflamación de estructuras musculoesqueléticas. Se trata de una enfermedad muy limitante, que produce dolor constante y disminuye tanto la calidad de vida como la funcionalidad física del paciente [1]. La inflamación crónica, hinchazón y rigidez en las articulaciones son sus principales síntomas, pero también puede afectar a órganos vitales, así como predisponer a los pacientes a padecer osteoporosis, enfermedades cardiovasculares, infecciones graves e incluso cáncer [2].

Los enfermos de artritis reumatoide pueden presentar inflamación en varias articulaciones, denominándose poliartritis, en dos o tres articulaciones siendo oligoartritis o solamente en una, llamándose monoartritis. La poliartritis es la presentación más común, siendo la monoartritis atípica.

Algunos pacientes, sin manifestar inflamación articular evidente, pueden desarrollar pericarditis, pleuritis, bronquiectasia o enfermedad intersticial pulmonar (EPI) [1].

La enfermedad suele aparecer entre los 40-60 años, pero puede afectar a personas de todas las edades, incluyendo a niños, denominándose artritis reumatoide juvenil [3]. La figura 1.1 muestra la prevalencia en grupos de edades estudiados en Colombia [4], también hay que añadir que existe una predisposición en cuanto al género, siendo las mujeres afectadas principalmente [5].

Tabla 1 – Prevalencia de artritis reumatoide por grupo etario en Colombia entre los años 2012 y 2016	
Grupo etario	Prevalencia%
De 0 a 4	0,02
De 5 a 9	0,05
De 10 a 14	0,11
De 15 a 19	0,14
De 20 a 24	0,14
De 25 a 29	0,22
De 30 a 34	0,38
De 35 a 39	0,55
De 40 a 44	0,79
De 45 a 49	1,12
De 50 a 54	1,52
De 55 a 59	1,75
De 60 a 64	1,88
De 65 a 69	1,96
De 70 a 74	1,97
De 75 a 79	1,86
De 80 o más	1,58
Total en mayores de 18 años	0,52

La prevalencia se calcula dividiendo el total de pacientes por grupo etario entre población media del periodo por grupo etario como denominador x 100 habitantes.

Figura 1.1: Riesgo de padecer artritis por rango de edades en Colombia.

1.3 Prevalencia

Actualmente, se estima que del 0,5-1% de la población padece como podemos apreciar en la **figura 1.2** artritis reumatoide [2].

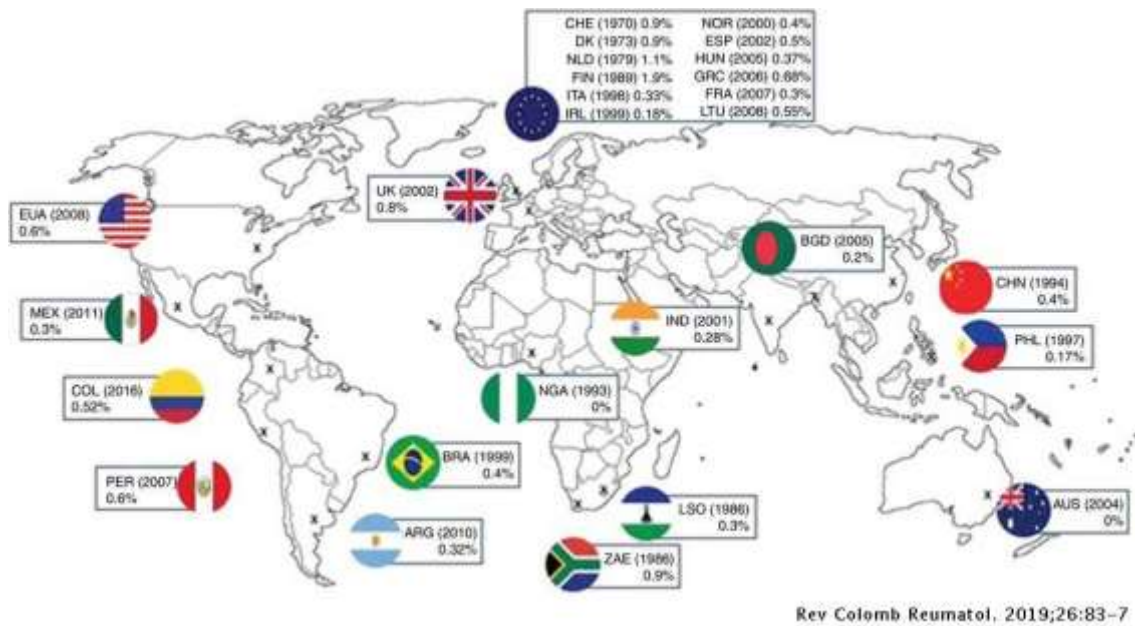


Figura 1.2: Resumen de datos de prevalencia por todo el mundo.

1.4 Aprendizaje automático

Como se comenta en el apartado 2, en la actualidad todavía existe la incertidumbre de saber que efectos tendrá combinar múltiples fármacos a corto, medio o largo plazo en pacientes con enfermedades crónicas. Por ello al no utilizarse actualmente ningún tipo de herramienta informática de *machine learning* en este tipo de pacientes, permite diseñar un algoritmo basándose en múltiples opciones aprendiendo de los datos proporcionados, prediciendo así los valores hepáticos en un tiempo determinado.

El aprendizaje automático o *machine learning* está contenido en la rama informática de Inteligencia Artificial (IA) y es una de las bases del proyecto, ya que ofrece la capacidad de resolver el problema mediante la adaptación de un modelo matemático a partir de datos reales, haciendo que el algoritmo aprenda, se adapte a los cambios y proporcione los resultados teniendo en cuenta todas las variables que previamente se han seleccionado. Estas variables utilizadas en el proyecto han sido seleccionadas en base a las evidencias del mundo real (RWE) y artículos científicos como explicamos en el apartado 3.2.

1.5 Diagnóstico

El diagnóstico de la artritis reumatoide se basa en los signos clínicos y pruebas de laboratorio. Los pacientes suelen acudir a consulta con inflamación articular o poliarticular simétrica e hinchazón. Las zonas afectadas más evidentes son metacarpofalángicas e interfalángicas proximales de las manos. También puede observarse inflamación de pulgares, muñecas y articulaciones metatarsofalángicas de los dedos de los pies.

En el examen físico, se realiza la prueba de compresión, que consiste en apretar los nudillos de manos y pies a nivel de las articulaciones metacarpofalángicas o metatarsofalángicas para comprobar la existencia de dolor. Es una herramienta útil pero por sí sola no es diagnóstica [6].

Los análisis de laboratorio más comunes para diagnosticar artritis reumatoide son:

- **Determinación de Factor Reumatoide (FR):** El factor reumatoide es un autoanticuerpo perteneciente al grupo de las inmunoglobulinas M (IgM), que se produce contra una fracción de inmunoglobulina G (IgG). Es una prueba que se realiza a pacientes con sospecha de artritis reumatoide, pero el FR también se encuentra elevado en personas con otras enfermedades crónicas e incluso en personas sanas [7]. Un resultado positivo no indica presencia de artritis reumatoide, pero un resultado negativo tampoco puede descartarla [8].
- **Anticuerpos antipéptidos cíclicos citrulinados (anti-CCP):** Son autoanticuerpos que se dirigen hacia una proteína o más del propio individuo. Este biomarcador junto al factor reumatoide FR es el que más se utiliza [9].
- Tanto el FR como los anti-CCP se consideraron biomarcadores séricos involucrados en la artritis reumatoide en el 2010, por el Colegio Americano de Reumatología (ACR) y la Liga Europea Contra el Reumatismo (EULAR). Posteriormente, se investigaron nuevos biomarcadores más sensibles y específicos para el diagnóstico de la artritis reumatoide [6].
- **Anticuerpos Antivimentina Citrulinada (anti-MVC):** son biomarcadores específicos que se emplean para el diagnóstico temprano de la enfermedad. Puede resultar positivo sin que el factor reumatoide se encuentre alterado. Aunque ha mostrado más sensibilidad y especificidad frente a los anti-CCP, se emplea como segunda línea para el diagnóstico de artritis reumatoide, quedando como primera opción el factor reumatoide y los anti-CCP [9].
- **Anticuerpos antinucleares (ANA):** Son anticuerpos creados por el sistema inmunitario que reaccionan frente a proteínas del mismo organismo. Son útiles en el diagnóstico de enfermedades autoinmunes [10].



- Proteína C Reactiva (PCR): Es una proteína que se eleva en la sangre en presencia de inflamación importante en el organismo [1].
- Velocidad de Sedimentación globular (VSG): Es la velocidad a la que descienden los glóbulos rojos en un tubo de sangre durante una hora. Se considera un factor de inflamación y se ve aumentado en enfermos de artritis reumatoide [11].
- Para valorar la actividad de la artritis reumatoide también se recurre a criterios como los del American College of Rheumatology (ACR), el Disease Activity Score (DAS) y el Simplified Disease Activity Index (SDAI) [1].
- El Disease Activity Score (DAS) que actualmente más se utiliza es el DAS28, que se basa en el recuento de 28 articulaciones dolorosas. Está recomendado por la EULAR [5].
- La identificación temprana de la enfermedad se ha relacionado con una disminución de lesiones articulares permanentes y una mejoría de la calidad de vida del paciente. Se están investigando nuevos biomarcadores más específicos de la enfermedad [6].

1.6 Tratamiento

El tratamiento de la artritis reumatoide consiste en terapias físicas, ayuda y educación del paciente y terapia farmacológica. Actualmente la artritis reumatoide no tiene cura, pero las terapias combinadas están dirigidas a producir una disminución de los síntomas clínicos, bajar la inflamación, reducir los daños irreversibles producidos por la inflamación crónica y prevenir la invalidez de los pacientes, mejorando su calidad de vida. [12].

En el momento de empezar el tratamiento, deben valorarse ciertos parámetros, para ello se realiza:

- Hemograma y bioquímica. Los parámetros bioquímicos relevantes son: Creatinina sérica, perfil hepático, velocidad de sedimentación globular (VSG) y proteína C reactiva (PCR) [13].
- Prueba de la hepatitis B (VHB) y C. Estas pruebas se hacen sobre todo antes de comenzar con el tratamiento de Leflunomida (LEF), fármacos biológicos y Metotrexato (MTX) [11].

Dentro del grupo de fármacos modificadores de la inflamación tenemos:

- Antiinflamatorios no esteroideos (AINES).

Son fármacos que actúan como coadyuvantes en el tratamiento de la artritis pero no modifican la enfermedad, por lo que no deben administrarse como única terapia [12]. Son antiinflamatorios, antipiréticos y analgésicos. Su diana principal es la ciclooxigenasa (COX) que es la enzima encargada de la biosíntesis de prostaglandinas y tromboxanos, fundamentales en la cascada de inflamación. Los AINES se asocian con toxicidad cardiovascular y problemas digestivos. En el tratamiento de la artritis reumatoide no se recomienda un AINE en concreto. La prescripción de estos fármacos depende de cada paciente [14].

En el tratamiento antiinflamatorio de la artritis reumatoide se encuentran a menudo los glucocorticoides [12]. Dependiendo de la dosis, los GC pueden tener acción antiinflamatoria o inmunosupresora. Tienen un metabolismo hepático y excreción renal [15]. En la artritis reumatoide los GC pueden administrarse vía intramuscular, oral e intraarticular [6]. Las terapias crónicas con GC tienen efectos secundarios graves que deben monitorizarse. Se ha demostrado que el empleo crónico de GC conlleva un aumento de triglicéridos en sangre y en algunos casos se ha notificado un aumento de colesterol. La hipertensión también es un efecto secundario ligado al uso de GC crónicos. El uso de GC junto a AINEs aumenta el riesgo de desarrollar problemas gastrointestinales, por lo que es fundamental la administración de protectores gástricos [16].

- Fármacos antirreumáticos modificadores de la enfermedad sintéticos (FAME)

Los FAME tienen como objetivo reducir los signos clínicos y lograr un grado bajo de actividad de la enfermedad.

Dentro del grupo de FAME sintéticos encontramos medicamentos como el Metotrexato, la Leflunomida, Sulfasalazina e Hidroxicloroquina [12].



Diseño de software para obtener la interacción toxicológica entre fármacos en pacientes con enfermedades crónicas

A continuación se describen el Metotrexato y la Leflunomida ya que es uno de los fármacos estudiados en este proyecto:

- El Metotrexato (MTX) es un medicamento que tiene buenos resultados para el control de la enfermedad en monoterapia o en combinación con otros fármacos. Reduce la mortalidad de los pacientes y es económico. Puede administrarse vía oral o vía parenteral, dependiendo de la tolerancia digestiva individual. Entre los efectos secundarios más destacados se encuentra: Mielosupresión, hepatotoxicidad, teratogenicidad, neumonía e infecciones [17].
- La Leflunomida es un derivado de isoxazol que tiene efectos antiproliferativos y antiinflamatorios [18]. Se trata de un medicamento que puede sustituir al MTX cuando éste no se tolera por parte del paciente. Los pacientes tratados con Leflunomida deben realizarse hemogramas y bioquímicas con monitorización de parámetros hepáticos ya que en diversos estudios se ha comprobado el efecto hepatotóxico de la Leflunomida [19].
- Fármacos antirreumáticos modificadores de la enfermedad biológicos:

Son anticuerpos humanos o proteínas de fusión que se dirigen contra citoquinas específicas proinflamatorias. Su acción es rápida y mejoran considerablemente la calidad de vida del paciente reduciendo la actividad inflamatoria de la enfermedad. Se dividen en 5 grupos en función de su mecanismo de acción [12]:

1. Antagonista de IL-6: Tocilizumab (TCZ).
2. Anticuerpo monoclonal anti-CD20: Rituximab (RTX).
3. Anticuerpos anti-TNF alfa: Infliximab (INF), Adalimumab, Golimumab (GOL), Etanercept (ETN) y Certolizumab (CTZ).
4. Proteína de fusión moduladora de la actividad de las células T: Abatacept (ABT).
5. Antagonista de la IL-1: Anakinra (ANK) [12].

A continuación se describe el Tocilizumab ya que es uno de los fármacos estudiados en este proyecto:

El Tocilizumab es un anticuerpo monoclonal IgG recombinante humanizado antirreceptor de IL-6 humana. Actualmente existen dos vías de administración, la intravenosa y la subcutánea. Puede utilizarse como terapia única o combinada con MTX. El Tocilizumab produce una elevación de las transaminasas hepáticas que debe monitorizarse con hemogramas y bioquímicas [20].

- Otros fármacos coadyuvantes en la terapia combinada de la artritis reumatoide son:

Fármacos antimitóticos:

Colchicina: Es un fármaco que actúa interfiriendo en la movilidad de leucocitos, fagocitando los cristales de urato y también tiene actividad antimitótica. En las crisis artríticas es común encontrar cristales de urato monosódico en los leucocitos del líquido sinovial, la Colchicina ayuda a controlar los cristales de urato. Entre los efectos secundarios que produce, los gastrointestinales dificultan en ocasiones su utilización, aunque también se ha estudiado el efecto mutagénico [21].

Fármacos inhibidores de la janus-quinasa:

Tofacitinib: Se trata de un inhibidor de la janus-quinasa, cuya acción se centra en inhibir la señalización y activación de inflamación en el organismo. La inhibición de la janus-quinasa (JAK) tiene un efecto inmunosupresor, que resulta de utilidad en el tratamiento de enfermedades autoinmune, como la artritis reumatoide. Entre los efectos secundarios adversos destaca el aumento de infecciones, ya que el sistema inmunitario se ve debilitado [22].



Analgésicos:

Debido a la inflamación y dolor articular, los pacientes con artritis reumatoide necesitan en numerosas ocasiones una terapia analgésica.

Hay distintos tipos de fármacos analgésicos que se emplean habitualmente en pacientes con artritis reumatoide. Pueden administrarse desde analgésicos para dolores leves hasta dolores moderados o severos [12].

El Paracetamol o Acetaminofén, es un fármaco analgésico de acción suave a moderada y antipirético. Puede encontrarse solo o combinado con otros fármacos, por ejemplo, combinado con opiáceos (Hidrocloruro de Tramadol), para el tratamiento de dolores severos [22]. Uno de los efectos nocivos más importantes es la necrosis hepática, que puede darse cuando la dosis diaria recomendada se sobrepasa [23].

Protectores gástricos:

Los fármacos antiinflamatorios tales como AINES y GC tienen un efecto inhibitor en la síntesis de prostaglandinas, que son fundamentales en la protección de la mucosa gástrica [24], [16]. Además, el empleo combinado de estos antiinflamatorios aumenta considerablemente la probabilidad de padecer complicaciones gastrointestinales como úlceras gástricas. Es por ello que los pacientes de artritis reumatoide tratados con antiinflamatorios necesitan protectores gástricos [12].

El Omeprazol es un protector gástrico conocido a nivel internacional. Es un inhibidor de la bomba de protones que disminuye la producción de ácido clorhídrico, disminuyendo así la acidez del estómago. Entre los efectos secundarios cabe destacar la posibilidad de alteraciones en analíticas correspondientes a problemas hepáticos [25].

La gran mayoría de fármacos que se utilizan hoy en día para el tratamiento de múltiples enfermedades y síntomas se metabolizan en el hígado. Es un órgano vital que posee numerosas funciones, desde almacenamiento de glucógeno hasta producción de enzimas fundamentales para la digestión, producción de colesterol y síntesis de factores de coagulación [26].

Se recomienda monitorizar el estado del hígado antes y después de aplicar terapias farmacológicas crónicas ya que, al ser el órgano implicado directamente en el metabolismo de muchos fármacos, lo hace vulnerable a la acumulación de sustancias tóxicas [27].

Para monitorizar el estado del hígado a nivel bioquímico se realiza la medición de enzimas como Alanina aminotransferasa (ALT), Aspartato aminotransferasa (AST), Gamma-glutamyl transferasa (GGT), Fosfatasa alcalina (FAL). También se controlan los parámetros de bilirrubina sérica, albúmina triglicéridos y colesterol, que son sustancias producidas en el hígado, así como la actividad de la protrombina [28].

Los enfermos de artritis reumatoide son tratados con fármacos que elevan las enzimas hepáticas y pueden producir hepatotoxicidad [27].

Estos fármacos son, fundamentalmente:

- Tocilizumab [29], Metotrexato [17], Paracetamol [23], Leflunomida [19], Omeprazol [25] y AINES [30].

El alcohol [31] y tabaquismo [32] suponen un riesgo añadido de hepatotoxicidad en enfermos de artritis reumatoide, por lo que es frecuente encontrar valores hepáticos alterados en las analíticas.

1.7 Hipótesis

El presente trabajo final de grado tiene como hipótesis:

- Mediar a un paciente teniendo en cuenta una predicción sobre la toxicidad que causarán la combinación de fármacos en dicho paciente puede ser una herramienta muy útil para el experto.

1.8 Objetivos

El objetivo principal del presente trabajo es diseñar, elaborar y validar un software que prediga los niveles de toxicidad en relación al tratamiento farmacológico, teniendo en cuenta la futura interacción toxicológica y por lo tanto permitiendo tomar la mejor decisión para el paciente y mejorando así la calidad de vida del mismo. Para ello será necesario cumplir los siguientes objetivos:

1. Investigación del estado del arte sobre la interacción toxicológica.
2. Diseño de algoritmo/s para poder predecir los valores toxicológicos teniendo en cuenta el máximo de variables posibles para que esta predicción sea lo más exacta.
3. Entrenamiento y obtención de valores utilizando redes neuronales con múltiples variables.
4. Comparación del estado del arte con los resultados obtenidos.

1.9 Resumen de apartados restantes

- **Capítulo 2.** Estado del arte, en este apartado se estudian los diferentes modelos, métodos y algoritmos de la actualidad, además de como se recogen los datos usados en estos métodos actualmente.
- **Capítulo 3.** Análisis del problema, en dicho apartado se describen los materiales y herramientas utilizadas en este proyecto, así como la justificación ética y científica.
- **Capítulo 4.** Este apartado describe la metodología usada explicando por todas las fases que ha pasado este proyecto.
- **Capítulo 5.** En este punto se analizan las soluciones propuestas con sus respectivos resultados.
- **Capítulo 6.** Conclusiones del proyecto respecto a los resultados obtenidos.

- **Capítulo 7.** Descripción de la relación del trabajo desarrollado con los estudios cursados.
- **Capítulo 8.** Estudio de trabajos futuros con su planteamiento.



2. Estado del arte

En este capítulo realizaremos una revisión de los métodos actuales para detectar la toxicidad hepática, además de revisar información de la interacción toxicológica entre fármacos y los modelos y/o métodos actualmente latentes.

2.1 Contexto

La labor de obtener la interacción toxicológica entre fármacos está poco estudiada, y en pacientes con enfermedades crónicas bastante menos aún. Esta tarea no es sencilla por la gran variedad de datos que pueden influir en estos valores, y esto sin el añadido de que sea una enfermedad crónica, ya que no se puede usar el mismo modelo de predicción de interacción toxicológica, porque no tienen el mismo efecto el uso de los mismos fármacos en un paciente con artritis reumatoide que en otro paciente que tenga diabetes. Esto es debido a que algunas enfermedades crónicas afectan a esta interacción e incluso a los valores toxicológicos del paciente.

La principal dificultad de este proyecto es elegir las herramientas y métodos que más se adecúen, además de la elección y recopilación de los datos para la obtención de mejores resultados. Otro añadido que presenta una dificultad es la aparición constante de artículos científicos, los cuales dejan en evidencia a los anteriores en cortos periodos de tiempo.

2.2 Algoritmos de obtención toxicológica

Actualmente existen prevalencias sobre *machine learning* aplicada a la medicina, como modelos de diagnosis de enfermedades, modificación del comportamiento, desarrollo de nuevos fármacos y aplicables a la radiología entre muchos otros.

Como describe IBM Watson Genomics [33] en 2016 empezaron a usar machine learning en áreas con altas necesidades como la identificación de cáncer [34] utilizando modelos de *machine learning*.

Respecto a la modificación de comportamiento IBM Watson Genomics dispone de diferentes tipos de software específicos basados en *machine learning*, uno de sus proyectos se basa en detectar los movimientos de mano a boca con el fin de que el paciente deje de fumar. La misma empresa también tiene en proceso de desarrollo un software que permite enviar imágenes para preveer el riesgo de cáncer de piel, sigue en proceso ya que la precisión basada en la escala aún debe adaptarse.

Otro punto importante en el cual se utiliza *machine learning* es el descubrimiento y desarrollo de nuevos fármacos. La medicina de predicción clínica utiliza el mismo modelo con el fin de ajustar más este aprendizaje sobre la progresión de enfermedades, también se usa en predicciones multitarea de nuevas enfermedades a partir de pruebas de laboratorio [35].

En el ámbito de la radiología y la radioterapia se usa *machine learning* para acelerar el proceso de segmentación y aumentar la precisión en la planificación de la radioterapia [36] Google, DeepMind y UCLH son unas de las empresas que más están invirtiendo en estas nuevas necesidades [33].

Debido a lo comentado anteriormente, no existen algoritmos o métodos específicos para obtener la interacción toxicológica en enfermedades crónicas a corto o largo plazo. Esto no quiere decir que no haya herramientas actualmente cuya función sea predecir la toxicidad, como por ejemplo los modelos de predicción basados en varios métodos QSAR [37]. Este tipo de modelos se ha estudiado con la finalidad de predecir toxicidades que se vuelvan crónicas, aparentemente es útil para el presente proyecto pero la finalidad de este es muy diferente, ya que no se busca saber si x toxicidades van a ser crónicas.

2.3 Real World Evidence

El mundo está cambiando, así como el ser humano gestiona la información más relevante para sacarle el máximo partido, como detallan expertos de este ámbito en [38] se puede afirmar que



Diseño de software para obtener la interacción toxicológica entre fármacos en pacientes con enfermedades crónicas

estamos ante una nueva era, en la cual los datos a partir del RWE permite analizar y obtener valores de la práctica real, no como un estudio clínico, sino viendo la afección real de esta toma de medicamentos en millones de personas.

3. Análisis del problema

En este capítulo se expondrán tanto materiales como herramientas utilizadas para la realización del trabajo, empezando con la base de datos, como hemos recabado dichos datos, el tipo de medicaciones que vamos a estudiar y las herramientas necesarias para la realización del trabajo.

3.1 Marco legal y ético

Para poder realizar este proyecto se ha tenido que pasar por el comité ético de la UPV para que apruebe el proyecto. Para la aprobación de este proyecto se ha realizado una memoria compuesta por los apartados que se describen a continuación.

3.1.1 Justificación científica y ética.

En el proyecto se estudia la toxicidad de la interacción entre fármacos proporcionados durante el tratamiento de pacientes con enfermedades crónicas. En particular se analiza el nivel de toxicidad hepática causado por la interacción entre fármacos. Los resultados derivados de este proyecto serán de utilidad clínica, ya que facilitarán la prevención de dificultades desde el inicio de un tratamiento y durante el mismo, permitiendo una mejora de la calidad de vida del paciente, además de aportar información relevante a los clínicos que ayude en la mejor elección del tratamiento valorando las diferentes opciones existentes.

Teniendo en cuenta el interés de los resultados obtenidos en estudio previos relacionados con los niveles de toxicidad hepática, en este estudio se pretende buscar el tratamiento farmacológico que cause el mínimo daño en la salud del paciente. El objetivo específico del estudio es predecir los niveles de toxicidad hepática máximos que no causen un perjuicio en el paciente, valorando factores implicados como el consumo de alcohol, tabaco, sexo, edad, etc. proporcionando la máxima calidad de vida.

El estudio es observacional, por lo que en ningún caso el tratamiento y seguimiento habitual de los pacientes se verá afectado de ningún modo.

3.1.2 Metodología.

Para desarrollar el estudio propuesto, se recopilan datos de 130 pacientes tratados en la Asociación de Artritis Reumatoide de Valencia desde el 2013 hasta 2020.

Los datos facilitados por parte de la Asociación de Artritis Reumatoide consisten en los resultados de las analíticas realizadas a los pacientes totalmente anónimos. El grupo de investigación de la UPV en ningún caso recogerá datos demográficos ni números de registro del paciente. El dataset de trabajo que recoge el grupo de investigación BDSLab de la UPV consiste en una matriz numérica sin posible vinculación a los pacientes originales.

La metodología de trabajo del estudio propuesto consta de diferentes fases, que se recogen en la Tabla 3.1.2. En ella también se indica la institución responsable de llevar a cabo cada tarea.

Fases de trabajo	Tareas	Institución responsable
Fase 1	Recogida y anonimización de datos	Asociación de Artritis Reumatoide
Fase 2	Estudio de modelos matemáticos	Universidad Politécnica de Valencia
Fase 3	Desarrollo de algoritmos	Universidad Politécnica de Valencia
Fase 4	Implementación de algoritmos en programa	Universidad Politécnica de Valencia
Fase 5	Testing de programa	Universidad Politécnica de Valencia
Fase 6	Análisis de resultados	Universidad Politécnica de Valencia
Fase 7	Publicación de resultados obtenidos	Universidad Politécnica de Valencia

Tabla 3.1.2: Fases de trabajo, tareas e instituciones responsables de llevarlas a cabo.

3.1.3 Información que recibirán los participantes.

Todos los participantes han sido informados mediante el “Modelo de consentimiento informado del participante” en el que se indica:

- Los objetivos del proyecto y el grupo de investigación que lo desarrolla.
- El protocolo y las condiciones que se van a seguir durante el estudio.
- El tratamiento de los datos: anonimización por parte de la Asociación AVAAR y confidencialidad de los mismos.
- El uso que se le va a dar a los datos anónimos recogidos. Se les informa de que los datos serán tratados y custodiados según la vigente normativa de protección de datos y serán utilizados únicamente con el fin expuesto en el presente estudio.
- Del derecho de acceso, rectificación, cancelación y oposición que podrán ejercitar mediante solicitud ante el investigador responsable del procedimiento de gestión de datos

3.1.4 Procedimiento de gestión de los datos personales.

TRATAMIENTO DE DATOS

El tratamiento, comunicación y cesión de sus datos se hará conforme a lo dispuesto por el Real Decreto-ley 5/2018, de 27 de julio de 2018, de Protección de Datos de Carácter Personal. Sólo el equipo investigador tendrá acceso a todos los datos previamente anonimizados recogidos.

ANONIMIZACIÓN

Los datos recogidos por el grupo de investigación de la UPV son totalmente anónimos, sin posibilidad ni intención de identificar a los pacientes originales. En ningún caso se han solicitado datos identificativos del paciente o números de registro. El grupo de investigación recogerá matrices numéricas que no presentan posibilidad de ninguna vinculación con los pacientes originales.

3.1.5 Compensaciones económicas.

No se prevé compensación económica o de otro tipo para los participantes de la actividad. El investigador no recibirá retribución específica por la dedicación al estudio.

3.2 Conjunto de datos

Dicho conjunto de datos ha servido para el entrenamiento del algoritmo y para la evaluación de los resultados, teniendo estos en cuenta para cerciorarnos de la eficacia del algoritmo. Los datos recabados han sido donados por la asociación de artritis reumatoide AVAAR de la forma especificada en los apartados anteriores.

El conjunto de datos a tener en cuenta ha sido extraído teniendo en cuenta el RWE. El Real World Evidence en contexto medicinal significa que los datos obtenidos tienen una evidencia en el mundo real, fuera de estudios clínicos. Estos datos han sido generados gracias a la práctica clínica habitual como se observa en la imagen 1, esta imagen simula múltiples factores que afectan directa/indirectamente en los datos de los pacientes.

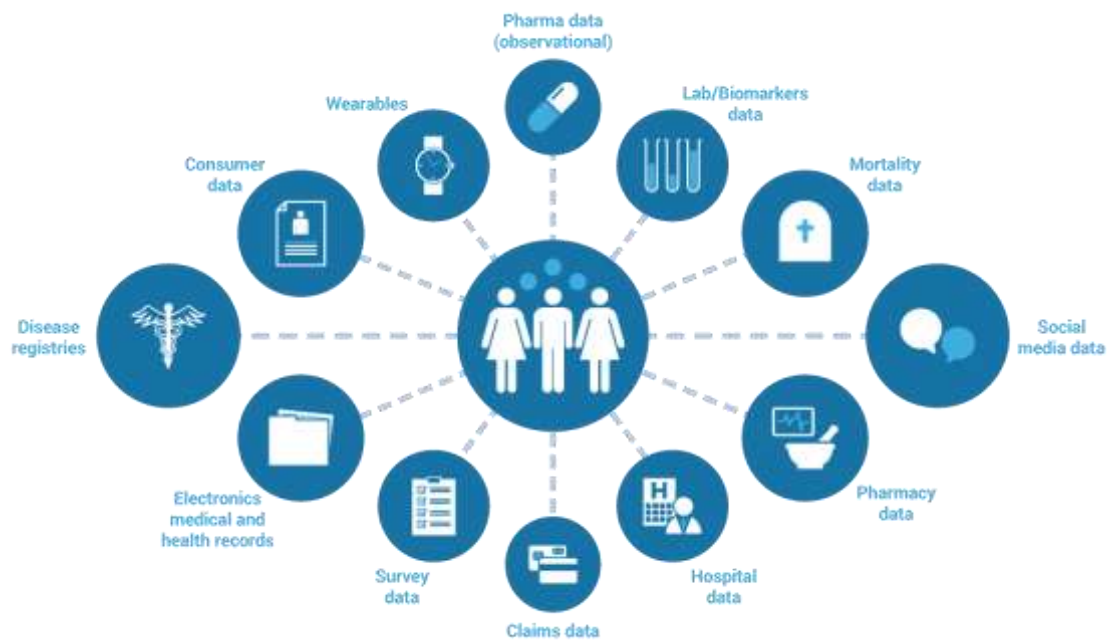


Figura 3.2: Real word evidence.

3.2.1 Descripción de los datos

Los datos utilizados en este estudio han sido proporcionados de forma totalmente anónima en forma de analíticas, se han utilizado dos analíticas por paciente, una de las analíticas con un año de diferencia antes de empezar con el tratamiento y la otra un año después. Además de estas analíticas también se ha tenido en cuenta la pauta de medicamentos desde el inicio de dichas analíticas y durante ese año de intervalo entre estas.

Respecto a los datos exactos de las analíticas, nos basamos concretamente en valores hepáticos como la alanina transferasa, aspartato transferasa y gamma glutamil transferasa, estos son los valores a tener en cuenta cuando hablamos de toxicidad hepática. Otros valores no hepáticos como los triglicéridos y el colesterol también son relevantes, ya que hay estudios que relacionan la toma del Tocilizumab, un medicamento tomado habitualmente en pacientes con esta enfermedad, con el aumento considerable de colesterol, el cual presenta un riesgo para su salud.

De todo el conjunto de datos recibidos han sido de utilidad una parte de estos datos proporcionados por la asociación de artritis reumatoide, ya que algunas analíticas que se han proporcionado no contenían la información suficiente, clínicamente hablando. También se han retirado del estudio los pacientes que solo han aportado una analítica, ya que sin una analítica previa y otra posterior no es viable para desarrollar un modelo predictor, es requisito indispensable.

Otro conjunto de datos utilizado es una matriz de pacientes x medicamentos, asignando con valores de 0/1 si x paciente toma ese medicamento (1), o si no lo toma (0). En este conjunto se hallan todos los medicamentos que toman los pacientes. Además de estos datos se han tenido en cuenta datos relevantes que afectan directa e indirectamente a la toxicidad hepática como si el paciente es fumador, ingiere alcohol, el sexo y la edad. Teniendo en cuenta este cúmulo de datos disponemos de 20 variables de entrada para su entrenamiento.

Los datos del estudio los vamos a mostrar en 3 tablas debido a su amplitud, en la tabla 3.2.2 disponemos de los datos relevantes no médicos con unos ejemplos de datos reales.

Id paciente	Edad	Sexo	Alcohol	Tabaco
1	53	1	0	0
2	61	1	0	0
3	46	1	1	0
4	53	0	0	0
5	64	0	1	0

Tabla 3.2.2: datos reales de pacientes.

donde el Sexo se representa con “1” para mujeres y con “0” para hombres, del mismo modo con para el tabaco y el alcohol, los cuales se representan con “1” si se fuma o se bebe. En cambio la edad es un valor real a tener en cuenta porque tiene un efecto directo con la degradación y el empeoramiento de nuestras funciones motoras, así como en órganos.

Los medicamentos también forman parte de este conjunto de datos de entrada a analizar, ya que de esta forma, dependiendo de los que tome cada paciente sus valores hepáticos, de colesterol y triglicéridos se verán afectados.

Medicamentos	Paciente 1	Paciente 2	Paciente 3
Naproxeno	1	1	1
Paracetamol	0	1	0
Dacortín	0	0	1
Xeljanz	0	0	0
Omeprazol	1	1	1
Arava	1	0	1
Tocilizumab	1	0	1
Colchicina	1	1	1
Metotrexato	0	1	0
Zaldiar	1	0	1

Tabla 3.2.3: datos reales de pacientes con sus respectivos medicamentos.



Los datos de la tabla 3.2 representados con un “1” significa que toman ese medicamento y un “0” que no lo toman. Los datos por sí son suficientes, pero al tener las variables de edad, sexo, alcohol y tabaco podemos ser más precisos en nuestra predicción toxicológica.

Teniendo en cuenta todos los valores anteriores debemos introducir en nuestro conjunto de datos los valores que son relevantes, sacados de las analíticas 1 año antes. Este conjunto de datos serán los valores de salida a predecir representado en la tabla 3.3 donde las siglas AST son de alanino transferasa, ALT de Alanina aminotransferase y GGT gamma-glutamil transferasa.

Id paciente	AST	ALT	Colesterol	Triglicéridos	GGT
1	13	13	188	142	10
2	15	11	156	79	11
3	20	17	130	86	10

Tabla3.2.4: valores de analíticas.

Estos datos de la última tabla la hemos tenido en cuenta en el entramiento de datos y la predicción, aplicando porcentajes de dichos valores sobre los valores máximos que dictan los médicos como valores de riesgo.

La fórmula y las variables que hemos tenido en cuenta para calcular estos valores son:

- Colesterol máximo: valor máximo de colesterol a partir del cual se considera riesgo.
- Colesterol actual: valor actual del colesterol en el momento de la analítica.

$$\text{Colesterol actual} / \text{Colesterol máximo} = \% \text{ Colesterol}$$

Con esta fórmula obtenemos el % colesterol hasta riesgo que usaremos como entrenamiento, siendo riesgo $\% \text{ Colesterol} > 1$.

- AST máxima: valor máximo de la AST a partir del cual se considera riesgo.
- AST actual: valor actual de la AST en el momento de la analítica.

$$\text{AST actual} / \text{AST máximo} = \% \text{ AST}$$

Con esta fórmula obtenemos el % AST hasta riesgo que usaremos como entrenamiento, siendo riesgo $\% \text{ AST} > 1$.

- ALT máxima: valor máximo de la ALT a partir del cual se considera riesgo.
- ALT actual: valor actual de la ALT en el momento de la analítica.

$$\text{ALT actual} / \text{ALT máximo} = \% \text{ ALT}$$

Con esta fórmula obtenemos el % ALT hasta riesgo que usaremos como entrenamiento, siendo riesgo $\% \text{ ALT} > 1$.

- Triglicéridos máximos: valor máximo de los triglicéridos a partir del cual se considera riesgo.
- Triglicéridos actuales: valor actual de los triglicéridos en el momento de la analítica.

$$\text{Triglicéridos actuales} / \text{Triglicéridos máximos} = \% \text{ Triglicéridos}$$

Con esta fórmula obtenemos el % triglicéridos hasta riesgo que usaremos como entrenamiento, siendo riesgo $\% \text{ triglicéridos} > 1$.

GGT máxima: valor máximo de la GGT a partir del cual se considera riesgo.

GGT actual: valor actual de la GGT en el momento de la analítica.

$$\text{GGT actual} / \text{GGT máximo} = \% \text{ GGT}$$

Con esta fórmula obtenemos el % GGT hasta riesgo que usaremos como entrenamiento, siendo riesgo $\% \text{ GGT} > 1$.

Con lo que la tabla teniendo en cuenta las formulas anteriores para el entrenamiento de los datos es la siguiente:



Id paciente	AST	ALT	Colesterol	Triglicéridos	GGT
1	0,393939394	0,40625	0,94	0,713	0,277
2	0,454545455	0,34375	0,78	0,396	0,305
3	0,606060606	0,53125	0,65	0,432	0,27

Tabla 3.2.5: valores de analíticas en % a predecir por nuestro programa.

De la misma forma que en la tabla 3.2.4 con la analítica posterior a 1 año obtenemos los valores de riesgo.

3.3 Equipamiento hardware

El hardware utilizado para el entrenamiento de los datos ha sido un ordenador portátil MSI GF63 Thin 95C con las siguientes características:

- Gráfica NVIDIA GTX 1650 MAX Q detallada en la tabla 3.3.1

Potencia de cómputo	2.55 TFLOPS
Frecuencia base	1020 MHz
Memoria	4 GB

Tabla 3.3.1: especificaciones técnicas de la tarjeta gráfica.

- Procesador Intel® Core™ i7-10750H descrito en tabla 3.3.2.

Velocidad del procesador	2.60bGHz
Número de núcleos	6

Velocidad del turbo procesador	5 GHz
---------------------------------------	-------

Tabla 3.3.2: especificaciones técnicas del procesador.

- Memoria RAM DDR4 16 GB.
- Disco SSD NVMe 256 GB.

3.4 Equipamiento Software

En la realización del presente proyecto se ha utilizado como herramienta principal de desarrollo Python v 3.9 con IDE PyCharm.

Para la construcción de las redes neuronales se ha utilizado la librería TensorFlow, una librería de código abierto y aprendizaje automático desarrollada por Google Brain. Más información acerca de esta librería en <https://www.tensorflow.org>.



4. Métodos

En el presente capítulo se describe la metodología utilizada para la realización del presente proyecto, poniendo énfasis en todas las fases por la que ha pasado este proyecto para la elaboración de un sistema predictor mediante redes neuronales.

4.1 Arquitecturas de redes neuronales y diseño detallado

Como hemos comentado anteriormente el proyecto consiste en obtener la interacción toxicológica entre fármacos utilizando redes neuronales. Para ello con las variables seleccionadas especificadas en el capítulo 3, se han aplicado diferentes arquitecturas de redes neuronales para comparar posteriormente los resultados y elegir el más eficiente.

Se ha estudiado el modelo de red neuronal utilizando capas intermedias ocultas y directamente con dos capas, una de entrada y otra de salida.

- **Red neuronal simple**

Este tipo de red consta de una capa de datos de entrada y otra capa de datos de salida, cuyos datos de entrada son los valores hepáticos, medicamentos y variables de edad, sexo, alcohol, etc. como se ha explicado en el apartado 3.

En la imagen 4.1 se observa unos ejemplos de datos de entrenamiento de entrada y en la imagen 4.2 los datos entrenamiento de salida.

```

[55, 1, 0, 0, 0.393939394, 0.46625, 0.94, 0.71557919, 0.277777778, 1, 0, 0, 0, 1, 1, 1, 1, 0, 1],
[61, 1, 0, 0, 0.575757576, 0.65625, 0.825, 0.582914571, 0.388888889, 1, 1, 0, 0, 1, 0, 0, 1, 0, 0],
[64, 1, 0, 0, 0.454545455, 0.34375, 0.78, 0.396984925, 0.305555556, 1, 0, 1, 0, 1, 1, 1, 1, 0, 1],
[58, 1, 0, 0, 0.469046904, 0.53125, 0.85, 0.432188804, 0.277777778, 1, 0, 0, 1, 1, 1, 1, 1, 0, 1],
[46, 1, 1, 0, 0.545454545, 0.46875, 0.78, 0.376884422, 0.222222222, 1, 1, 0, 0, 1, 1, 1, 0, 0, 0],
[53, 1, 0, 0, 0.454545455, 0.375, 0.935, 0.422110553, 0.388888889, 1, 0, 0, 0, 1, 1, 1, 1, 0, 1],
[64, 0, 1, 0, 0.518181818, 0.46875, 0.83, 0.462311558, 0.333333333, 1, 0, 0, 0, 1, 1, 1, 1, 0, 1],
[55, 0, 0, 0, 0.545454545, 0.71875, 0.99, 0.417085427, 0.305555556, 1, 1, 0, 0, 1, 0, 0, 1, 0, 0],
[62, 1, 0, 0, 0.393939394, 0.3125, 0.775, 0.621115578, 0.25, 1, 1, 0, 1, 1, 1, 1, 0, 0, 1],
[47, 1, 0, 0, 0.242424242, 0.21875, 1.01, 0.542713568, 0.388888889, 1, 1, 0, 0, 1, 0, 0, 1, 0, 0],
[49, 0, 0, 0, 0.303030303, 0.46625, 1.07, 0.628140704, 0.222222222, 1, 0, 0, 0, 1, 1, 1, 1, 0, 1],
[48, 0, 1, 0, 0.545454545, 0.5, 0.92, 0.5728, 0.361111111, 1, 0, 0, 0, 1, 1, 1, 1, 0, 1],

```

Figura 4.1: datos de entrada

Las cinco primeras variables de a Imagen 4.1 se corresponden en orden a edad, sexo, alcohol y tabaco, siguiendo con las cinco siguientes con los valores toxicológicos en orden: alanino transferasa, alanina aminotransferasa, colesterol, triglicéridos y gamma-glutamyl transferasa. Siguiendo con el resto de variables, las 10 restantes se refieren al tipo de medicación tomada, se enuncian en orden: Naproxeno, Paracetamol, Dacortín, Xeljanz, Omeprazol, Arava, Tocilizumab, Colchicina, Metrotrexato y Zaldiar

```

[0.484848485, 0.46875, 1.385, 0.592964824, 0.472222222],
[0.545454545, 0.78125, 1.09, 0.628140704, 0.388888889],
[0.545454545, 0.4375, 1.16, 0.633165829, 0.388888889],
[0.757575758, 0.5625, 0.72, 0.964824121, 0.361111111],
[0.575757576, 0.71875, 0.83, 0.462311558, 0.444444444],
[0.363636364, 0.53125, 1.32, 0.40201005, 0.444444444],
[0.606060606, 0.6875, 1.16, 0.708542714, 0.527777778],
[0.96969697, 1.375, 1.305, 0.638190955, 0.777777778],

```

Figura 4.2: datos de salida

Respecto a los datos de entrenamiento de salida, estas variables han sido extraídas de los resultados un año después de su anterior analítica, viendo así el efecto dañino de estas interacciones, así mismo el orden de estas variables es: alanino transferasa, alanina aminotransferasa, colesterol, triglicéridos y gamma-glutamyl transferasa. Todos estos datos están formateados según el capítulo 3 para que se indique con un valor >1 para los valores en los que se deben contemplar como riesgo.

En la imagen 4.3 se muestra como se define el modelo a utilizar, como se observa solo tiene la capa de entrada y de salida, cinco datos de salida y diecinueve de entrada.



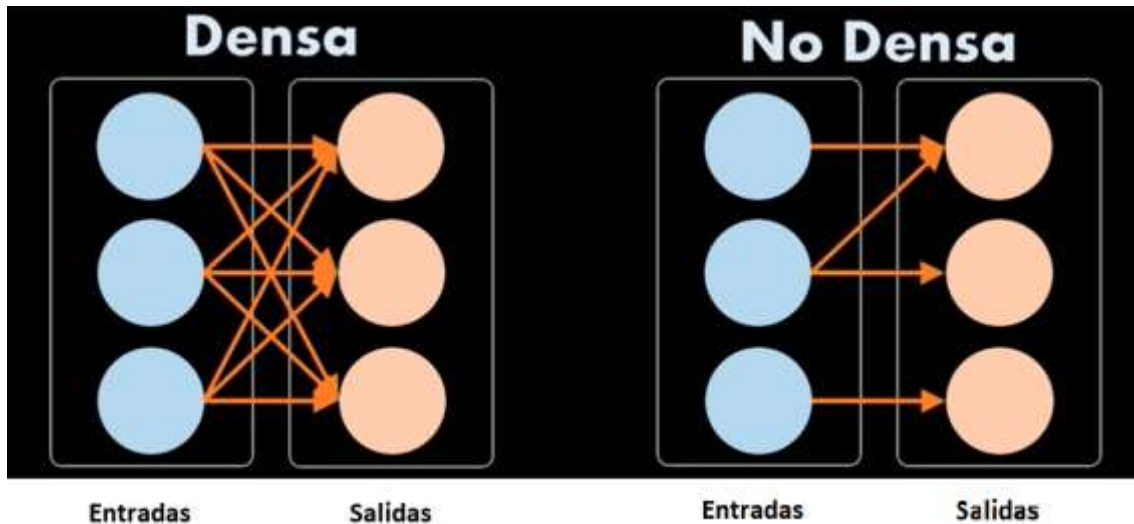


Imagen 4.3: arquitectura de red neuronal densa y no densa.

Además se ha utilizado un modelo secuencial con una la interacción entre neuronas *Dense*, ya que este tipo de capas nos dan más opciones al relacionarse todas las neuronas de entrada con todas las de salida como se observa en la imagen 4.4. Por otro lado, se ha utilizado optimizador Adam que se basa en la estimación adaptativa, la tasa de aprendizaje (e.g 0.01) para ajustar los pesos.

Para calcular el error, en este caso se ha optado también por el tipo *mean_squared_error* el cual quiere decir que muchos errores pequeños da un mejor resultado que pocos errores grandes.

```
modelo = tf.keras.models.Sequential()

modelo.add(tf.keras.layers.Dense(5, input_dim=19))

modelo.compile(
    optimizer=tf.keras.optimizers.Adam(0.01),
    loss='mean_squared_error'
)
```

Figura 4.4: código del modelo.

Volviendo al modelo de red neuronal utilizado, se ha buscado la cantidad de iteraciones de aprendizaje llamadas *epochs* que podemos visualizar en la imagen 4.5 a partir de las cuales se obtenga un resultado cercano al real. Para ello hemos entrenado el modelo con varios valores de iteraciones hasta encontrar el punto en el que se estabilizan los resultados.

```
print("Inicio de entrenamiento de datos")
historial = modelo.fit(inp, out, epochs=4000, verbose=False)
print("Modelo entrenado\n")
```

Figura 4.5: entrenamiento de datos.

En la imagen 4.5 se observa que con el método `.fit()` se entrenan las entradas y salidas denominadas en este caso como `inp` y `out` respectivamente, seguidas del número de iteraciones explicado anteriormente. El último parámetro denominado *verbose* tiene la función de imprimir por pantalla el progreso de las épocas, al situarlo en *False* no muestra este progreso, ya que interesa el entrenamiento total.

En la imagen 4.6 se observa un ejemplo de tiempo hasta estabilización de la red neuronal usada con la arquitectura sin capas intermedias, solo de entrada y de salida de la red neuronal que se ha empleado en este proyecto con los datos mencionados anteriormente.

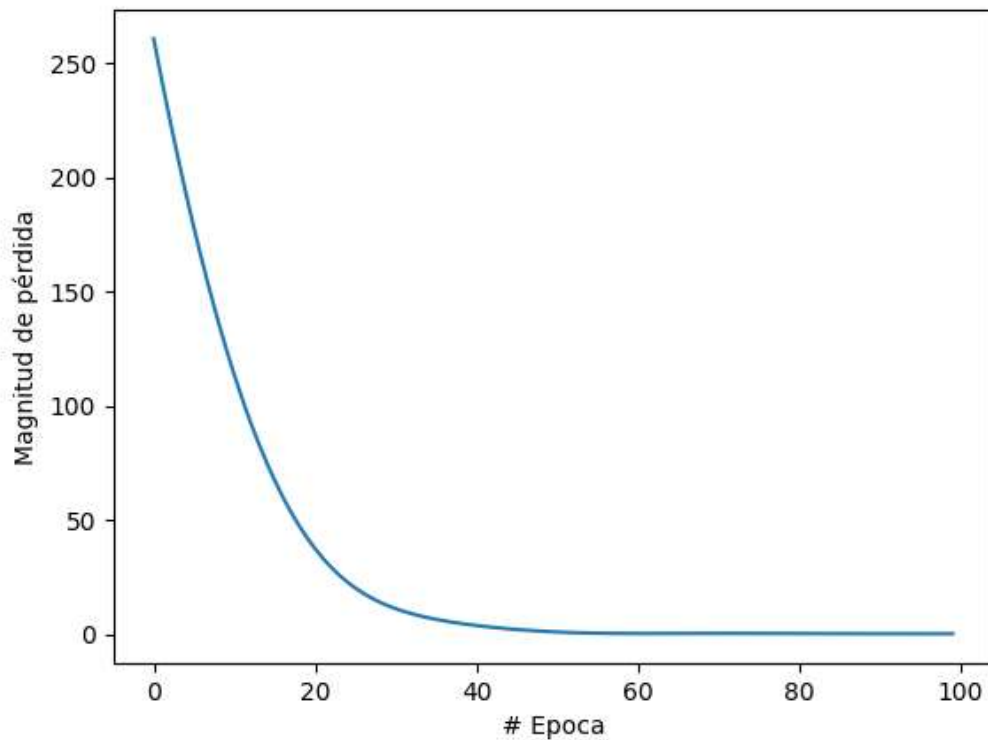


Figura 4.6: arquitectura sin capa, eje y: entrenamiento hasta estabilización, eje x: error.

- **Red neuronal con múltiples capas**

A diferencia de la red neuronal anterior este tipo de arquitectura no se basa en una entrada y una salida solamente, sino que entre la entrada y la salida hay por lo menos una capa oculta como se muestra en la imagen 4.7. El uso de estas capas es debido a la velocidad de aprendizaje, utilizar capas ocultas hace que se establezcan los resultados en menos iteraciones de entrenamiento, pero compromete el tiempo de entrenamiento.

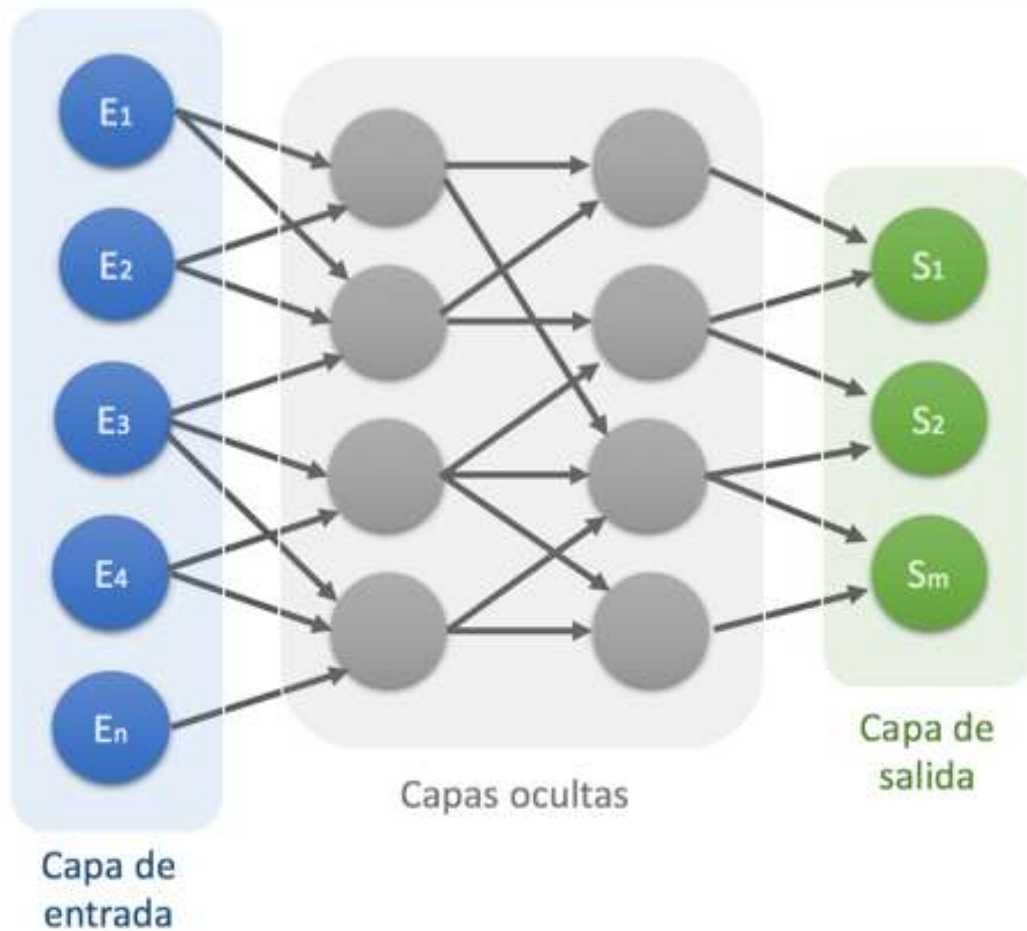


Figura 4.7: ejemplo de arquitectura de red neuronal con capas ocultas.

Para visualizar al tiempo de estabilización se han hecho pruebas como en la anterior arquitectura, teniendo en cuenta con los mismos datos de entrada y el mismo tipo de error, número de *epochs*, etc. en este caso la imagen 4.8 muestra la utilización de una capa oculta intermedia en la red neuronal del proyecto y la imagen 4.9 la arquitectura de este tipo de red neuronal.

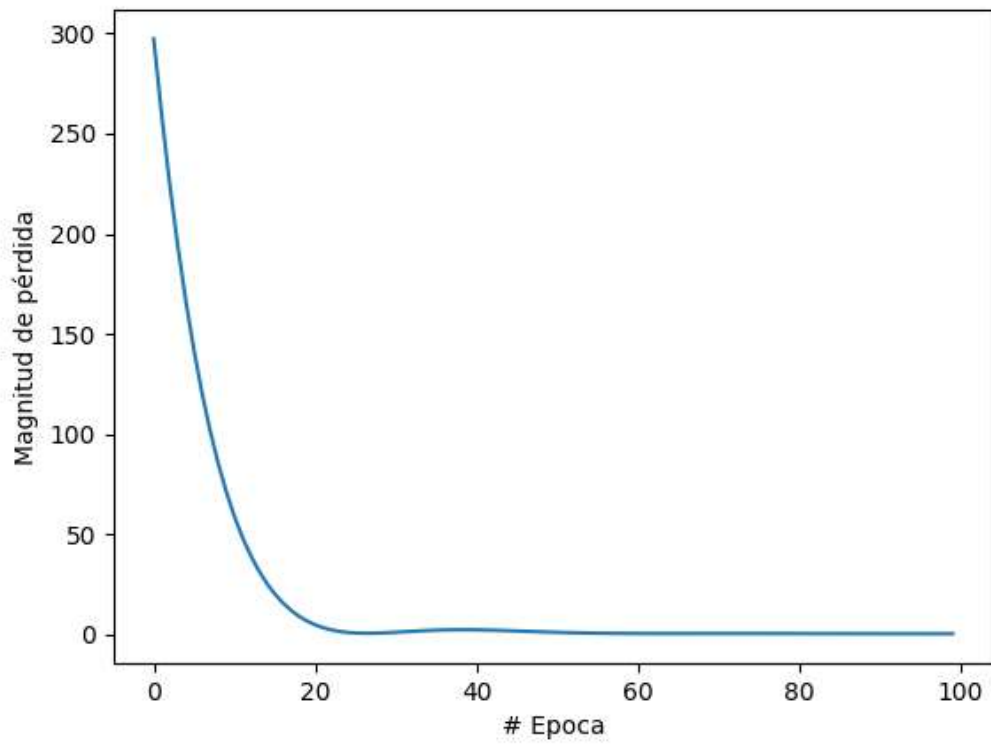


Figura 4.8: arquitectura con una capa oculta, eje y: entrenamiento hasta estabilización, eje x: error.

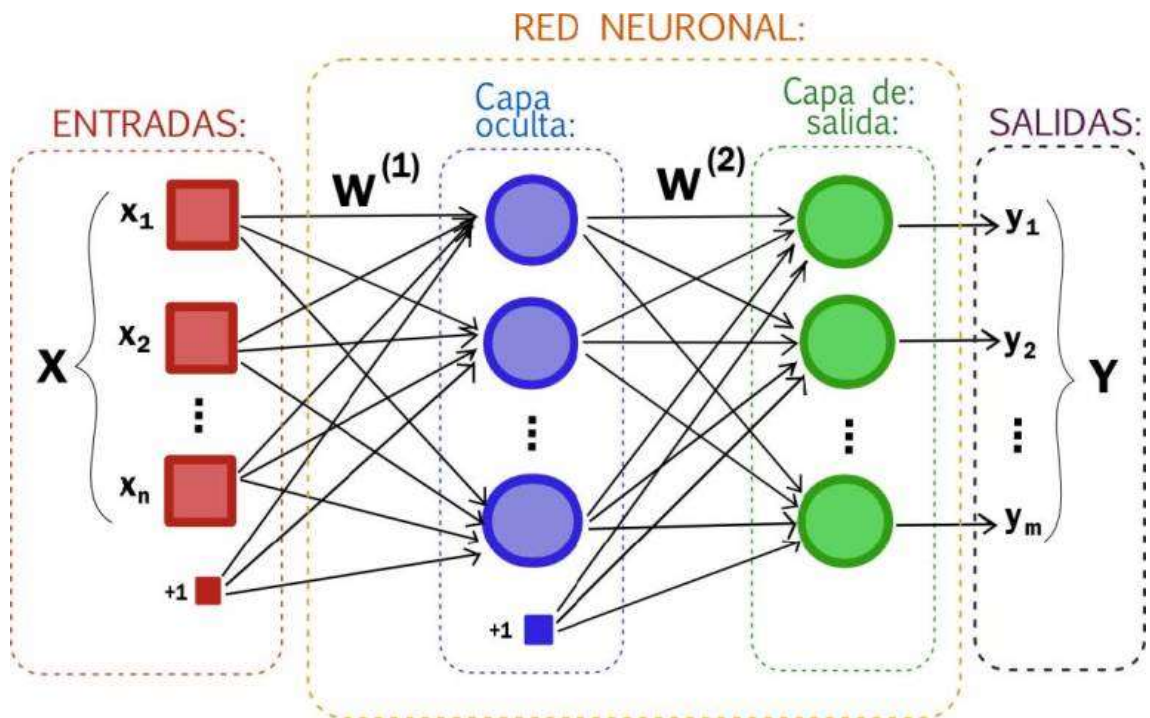


Figura 4.9: modelo de red neuronal con capa oculta intermedia.

Con la arquitectura de dos capas ocultas aplicadas en nuestro proyecto se muestra en la imagen 4.10 se obtiene la siguiente gráfica de aprendizaje:

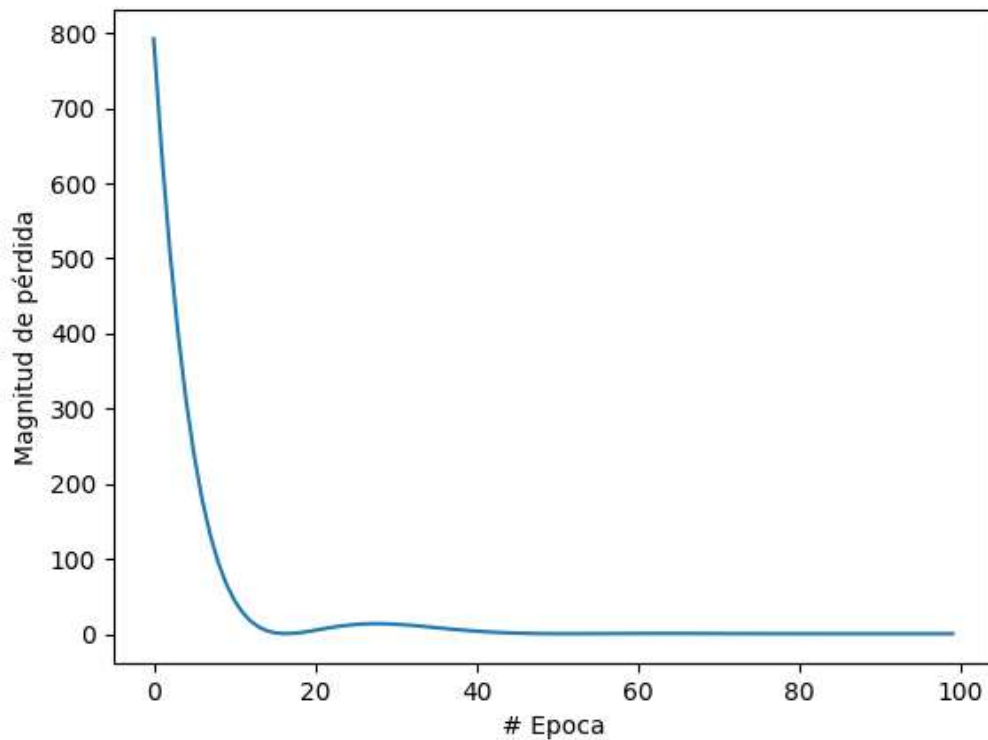


Figura 4.10: arquitectura con dos capas ocultas, eje y: entrenamiento hasta estabilización, eje x: error.

4.2 Evaluación

La evaluación de las etiquetas de salida obtenidas, para las cuales su obtención se ha explicado en el apartado 3.2.2, ha sido realizada a partir de factores de riesgo. En el momento en que un valor como el colesterol, la AST, ALT, GGT o triglicéridos es > 1 en la predicción de la red neuronal, se toma como valor de riesgo.

Valor de variable de salida $> 1 =$ riesgo.

En el siguiente apartado se explican detalladamente los resultados obtenidos.

5. Desarrollo de la solución propuesta y resultados

En este apartado se analizan los resultados obtenidos a partir de los distintos tipos de soluciones propuestas, así como una breve discusión de qué modelo de los estudiados es más apropiado para la finalidad del proyecto. Discutiéndose dichos resultados comparándolos respecto a la eficiencia y relevancia clínica.

5.1 Predicción toxicológica

En este subapartado se describen los resultados junto con los tiempos dependiendo del tipo de modelo utilizado.

5.1.1 Resultados con modelo simple

El modelo de red neural probado por primera vez ha sido el modelo simple denso, utilizando las capas de entrada y salida para el entrenamiento de datos, sin capas ocultas intermedias como se observa en la Imagen 4.3.

Este tipo de arquitectura es de las más simples referida al contexto de red neuronal, ya que necesita de datos de entrada y datos de salida, con lo cual el aprendizaje y el entrenamiento se focaliza en ambas capas.

Pasando a los tiempos de ejecución del algoritmo en esta arquitectura “simple”, se observa en la tabla 5.1 un tiempo de 4.6558 segundos con 4000 iteraciones.

<i>Variables de entrada</i>									
<i>Edad, sexo, alcohol, tabaco y valores de toxicidad</i>	53	1	0	0	0.394	0.406	0.94	0.71	0.27
<i>Variables de entrada tipo de medicación</i>	1	0	0	0	1	1	1	0	1

Tabla 5.2.1: Variables de entrada.

<i>Valores de Salida</i>	<i>AST</i>	<i>ALT</i>	<i>Colesterol</i>	<i>Triglicéridos</i>	<i>GGT</i>
<i>Valores de toxicidad</i>	0.4071	0.4982	1.30	0.72	0.295

Tabla 5.2.2: Valores toxicológicos de salida.

Analizando las tablas anteriores 5.1 y 5.2 se distinguen en ella diferentes datos de la predicción:

- Datos de entrada: son los datos que están explicados en el apartado 3.2.2.
- Salida: la salida corresponde a los datos explicados en el apartado 3.2.2, muestra la predicción toxicológica de los valores AST, ALT, colesterol, triglicéridos y GGT.

Como se explicará en el apartado de conclusiones no hay que basarse en los tiempos de ejecución o tiempos de estabilización, sino en la combinación de estos con la aproximación de la predicción a los resultados reales.

Siguiendo con esta solución, observamos que la cantidad de iteraciones hasta que estabiliza su aprendizaje es de unas 50 iteraciones, como se aprecia en la imagen 5.2

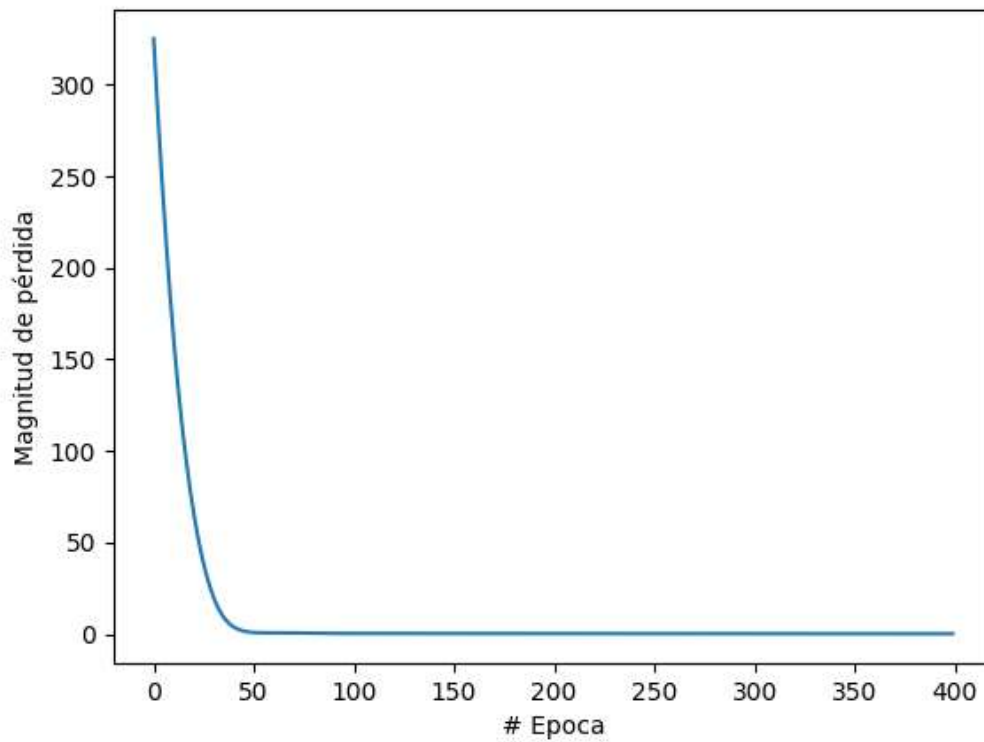


Figura 5.2: Iteraciones hasta estabilización de resultados. Eje y: entrenamiento hasta estabilización, eje x: error.

5.1.2 Resultados con modelo utilizando capas ocultas

En el caso de las soluciones implementadas utilizando capas ocultas, se observan dos tipos diferentes de impleteación, el modelo con una capa oculta intermedia y el otro modelo con dos capas ocultas intermedias.



▪ **Modelo con una capa oculta:**

En este caso se ha utilizado una capa intermedia oculta entre ambos modelos.

Con los mismos datos de entrada y en las mismas condiciones que en modelo simple explicado en el apartado anterior, con 4000 iteraciones, modelo *Dense*, etc. exceptuando el número de capas utilizadas, se han obtenido los siguientes resultados:

- Tiempo de ejecución: 5,09058 segundos.
- Iteración en la que se comienzan a estabilizar los resultados: 75 iteraciones aproximadamente como se observa en la figura 5.3.

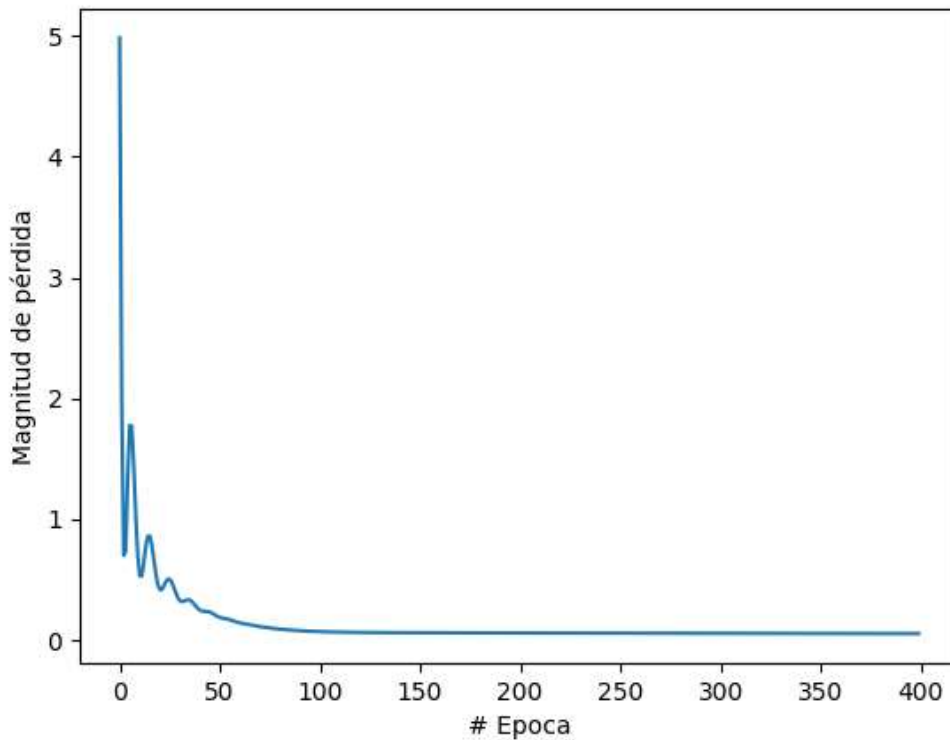


Figura 5.3: Iteraciones hasta estabilización con una capa oculta. Eje y: entrenamiento hasta estabilización, eje x: error.

- **Modelo con dos capas ocultas:**

La arquitectura de este modelo de red neuronal se distribuye de forma parecida al anterior, teniendo el añadido de tener otra capa más oculta.

Como se ha comentado en el modelo de una capa oculta para los cálculos de esta arquitectura se han utilizado los mismos datos de entrada, iteraciones, etc. solo cambia el añadido de otra capa oculta intermedia.

Resultados de este modelo:

- Tiempo de ejecución: 5,190983 segundos.
- Iteración en la que se comienzan a estabilizar los resultados: 70 iteraciones aproximadamente como se observa en la figura 5.4

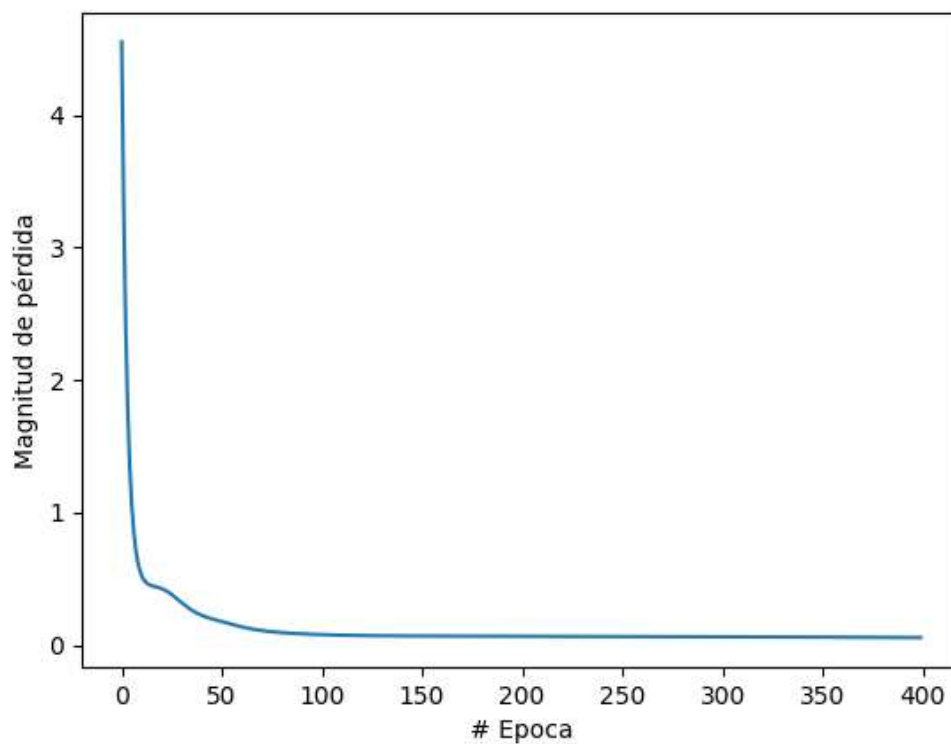


Figura 5.4: Iteraciones hasta estabilización con dos capas ocultas. Eje y: entrenamiento hasta estabilización, eje x: error.

5.2 Análisis de resultados

Como se ha detallado en el apartado 5, hay varias soluciones propuestas las cuales son válidas aparentemente, pero entre ellas las predicciones de resultados han sido considerablemente diferentes. En este apartado se va a explicar cual es el modelo que se aproxima más a los resultados reales.

Para decidir que modelo es más adecuado para la finalidad del proyecto han tenido en cuenta valores reales antes de hacer la predicción, mostrados en la figura 5.5, con los que compararemos los resultados de los modelos.

AST	ALT	Colesterol	Triglicéridos	GGT
[0.484848485,	0.46875,	1.385,	0.592964824,	0.472222222],

Figura 5.5: resultados reales de un paciente después de un año de tratamiento.

Se pueden observar diferencias mínimas entre los diferentes modelos en los resultados en las imágenes 5.6, 5.7 y 5.8, según el uso de una arquitectura u otra, respecto a tiempos de ejecución del programa o en iteraciones hasta una estabilización de las salidas. Estos resultados son relevantes en la ecuación, pero no los podemos tener en cuenta en la ecuación como un factor decisivo para elegir el modelo que mejor se ha adaptado a la finalidad del proyecto.

En el caso de los modelos que usan capas ocultas se ha comprobado que tienen un tiempo de ejecución mayor además de estabilizar las salidas en un tiempo mayor, este tiempo en el caso del presente proyecto no es muy relevante por los pocos datos de entrada que ha tenido, aproximadamente 400 datos, pero en un proyecto con más datos sí sería relevante, por ello lo hemos tenido en cuenta. Lo importante es que ninguno de estos modelos, al hacer una predicción de datos de los cuales ya se sabía la respuesta tienen un margen de la suma de sus errores de de un 10% el modelo con una capa y un 10,8% el modelo de dos capas ocultas. Los datos usando diferente número de capas se pueden ver en las figuras 5.6 y 5.7.

AST ALT Colesterol Triglicéridos GGT

```
[[0.47439516 0.41220936 1.0154989 0.6163091 0.44254118]]
```

Figura 5.6: datos de salida del modelo de red neuronal con una capa oculta.

AST ALT Colesterol Triglicéridos GGT

```
[[0.46418983 0.39204994 1.0148175 0.610656 0.45872247]]
```

Figura 5.7: datos de salida del modelo de red neuronal con dos capas ocultas.

En cambio con el modelo que no contiene ninguna capa oculta se han obtenido los mejores resultados, en cuanto a tiempo de ejecución y a la velocidad de estabilización de salidas. Basándonos en los resultados de salida reales, este modelo tiene un 83,2% de acierto, es decir un 16,8% de error en el computo total de las salidas.

En la siguiente figura 5.8 se muestran los datos de salida del modelo simple.

AST ALT Colesterol Triglicéridos GGT

```
[[0.4071084 0.49827152 1.3008864 0.7227227 0.29528973]]
```

Figura 5.8: datos de salida del modelo de red neuronal simple.

Concluyendo con los resultado se ha visto mayor tiempo y estabilización de los datos de salida en el modelo simple, pero este modelo tiene el mayor margen de error de los tres comparados descritos en la tabla 5.3.



Tipo de arquitectura de red neuronal	ALT	AST	Colesterol	Triglicéridos	GGT
Arquitectura simple	84%	95%	93%	82%	63%
Arquitectura con una capa oculta	97%	88%	73%	98%	94%
Arquitectura con dos capas ocultas	95%	93%	73%	98%	97%

Tabla 5.3: porcentaje de acierto de la predicción respecto a valor real.

- Error respecto a la predicción con datos reales en el modelo simple: 16,8%.
- Error respecto a la predicción con datos reales en el modelo con una capa oculta: 10%.
- Error respecto a la predicción con datos reales en el modelo con dos capas ocultas: 10,8%.

Estos datos indican que el modelo con una capa es el que más se acerca a los resultados de predicciones reales, pero no podemos escoger uno, dado que si nos fijamos en los valores de error individualmente, podemos observar que los valores menores de error en cada variable de salida han sido de diferentes modelos, es decir, podemos tener en cuenta los tres modelos y predecir un tipo de variable con cada uno basándonos en los siguientes resultados:

- AST: en la predicción de este valor el resultado más próximo ha sido mediante la red neuronal de una capa oculta, con un error de un 3%.
- ALT: en la predicción de este valor el resultado más próximo ha sido mediante la red neuronal simple, con un error de un 5%.
- Colesterol: en la predicción de este valor el resultado más próximo ha sido mediante la red neuronal simple, con un error de un 6%, lejos de las otras arquitecturas con un 23% de error en ambas.

- Triglicéridos: en la predicción de este valor el resultado más próximo ha sido mediante la red neuronal con dos capas ocultas, con un error de un 2%.
- GGT: en la predicción de este valor el resultado más próximo ha sido mediante la red neuronal con dos capas ocultas, con un error de un 3%.

Respecto a estos resultados, aclarar que se han entrenado cientos de veces en bloques de 4000 iteraciones, y no varían ni un 1% los datos de salida entre predicciones.

Con lo cual si escogemos un tipo de red para predecir las diferentes salidas, reducimos el error de la mejor predicción considerablemente. Bajamos este % de error del modelo que usa una capa oculta con un 10% de error a un 3,8% con un margen de + ó - 1% de error al combinar los diferentes modelos.



6. Conclusiones y reflexiones

En este apartado se exponen las conclusiones finales del presente proyecto.

6.1 Conclusiones

Evaluando los resultados descritos en el capítulo 5, se observa que el modelo de red neuronal con menos porcentaje de error es el modelo con la arquitectura de una capa oculta, seguido del modelo con dos capas ocultas. Los modelos que san capas ocultas tienen menos margen de error, pero algunos resultados son bastante inestables, en cambio del modelo simple, obtiene el mayor porcentaje de margen de error pero los resultados generalmente son más elocuentes, preciciendo riesgo donde debería haberlo y en caso contrario sin predecirlo.

Como se ha comentado en el capítulo 5, para según qué variable a predecir es más exacto un modelo u otro, así que para obtener el máximo grado de precisión se debe usar el modelo adecuado por cada variable a predecir.

6.2 Reflexión sobre el trabajo

Este presente trabajo ha superado una larga serie de problemas para ser resueltos. En primer lugar, se ha tenido que aprender durante el proyecto las bases del estudio científico, la programación en Python y del aprendizaje automático. Se han superado gracias a las capacidades adquiridas durante la carrera como Ingeniero.

Otro de los problemas principales ha sido la recopilación de los datos, ya que previamente para la aprobación de este proyecto se han tenido que firmar muchos documentos legales para la realización del presente proyecto, así como la aprobación de un comité ético. El problema de los datos ha estado principalmente en que ningún laboratorio nos ha abierto las puertas, y los hospitales que lo habían hecho debido a la situación de COVID no nos han podido atender. Como

resultado, se ha conseguido realizar el primer estudio de toxicología combinada de fármacos en pacientes con artritis reumatoide tras superar las fases de 1) análisis del problema y aprobación del comité ético, 2) recopilación de datos reales, 3) programación y evaluación de redes neuronales, 4) discusión científica de resultados.

7.Relación del trabajo con los estudios cursados

Como se ha comentado en el punto anterior, este trabajo tiene poca relación con la rama de ingeniería de computadores, la cual he cursado muy satisfactoriamente. Pese a no ser un trabajo propio de mi rama sí que ha habido una ayuda muy importante, dado que como optativa en último curso escogí Bioinformática, gracias a esta asignatura me decidí a realizar este proyecto.

Haber cursado Bioinformática me ha ayudado principalmente a plantear un proyecto con un impacto real en la sociedad y poder desarrollarlo, haciendo uso de las enseñanzas impartidas en esta asignatura para buscar información, dialogar con gente que padece este tipo de enfermedades, y con doctores que se dedican a tratarlas e investigar sobre estas.

Por ultimo pero no menos importante creo que la principal relación con los estudios cursados es que cada una de las asignaturas me ha ayudado directamente o indirectamente, así como haber superado algunas de ellas con muchísimo esfuerzo, y la capacidad de resolver problemas es algo que he ido aprendiendo en el transcurso de la carrera, cada vez mejor que la anterior.

8.Trabajos futuros

Respecto a los trabajos futuros relacionados con este proyecto, principalmente me hubiera gustado desarrollar una aplicación web Angular, de la cual se hicieron pruebas y se empezó a construir pero la complejidad de las redes neuronales, comenzar a trabajar en una empresa privada y el tiempo que ha pasado desde que nos proporcionaron los datos con los que hemos trabajado, ha desajustado el tiempo para la representación de este proyecto con una interfaz gráfica.

Añadido a lo anteriormente mencionado, respecto a la eficiencia me hubiera gustado poder probar con una cantidad de datos considerable como los de un estudio clínico, para así poder ser más preciso y acercarnos más a la realidad. También el *hardware* no ha supuesto un impedimento, pero con un equipo más orientado a este tipo de procesos se vería considerablemente más eficiente, sobre todo si pudiéramos optar a tantos datos como los que usa un estudio clínico. Por ultimo, otro camino a explorar sería indagar más sobre otro tipo de redes para compararlas con las utilizadas en este proyecto y seguir mejorando en nuestras predicciones.



9. Bibliografía

- [1] Sparks JA. Rheumatoid Arthritis. *Ann Intern Med.* 2019; 170(1):1-16.
- [2] van der Woude D, van der Helm-van Mil AHM. Update on the epidemiology, risk factors, and disease outcomes of rheumatoid arthritis. *Best Pract Res Clin Rheumatol.* 2018; 32(2):174-187.
- [3] Schneider R, Passo MH. Juvenile rheumatoid arthritis. *Rheum Dis Clin North Am.* 2002; (3):503-30.
- [4] Bernal-Macías S, Gutiérrez Dávila JM, Fernández-Ávila DG, Rincón-Riaño DN, Roselli D. Prevalencia de la artritis reumatoide en Colombia según información del Sistema Integral de Información de la Protección Social. *Rev Colomb Reumatol.* 2019; 26(2):83-87.
- [5] Allen A, Carville S, McKenna F; Guideline Development Group. Diagnosis and management of rheumatoid arthritis in adults: summary of updated NICE guidance. *BMJ.* 2018; 362:3015.
- [6] Littlejohn EA, Monrad SU. Early Diagnosis and Treatment of Rheumatoid Arthritis. *Prim Care.* 2018; 45(2):237-255.
- [7] Morgan GJ Jr, Chow WS. Clinical features, diagnosis, and prognosis in rheumatoid arthritis. *Curr Opin Rheumatol.* 1993; 5(2):184-90.

[8] Ingegnoli F, Castelli R, Gualtierotti R. Rheumatoid factors: clinical applications. *Dis Markers*. 2013; 35(6):727-34.

[9] Zhu JN, Nie LY, Lu XY, Wu HX. Meta-analysis: compared with anti-CCP and rheumatoid factor, could anti-MCV be the next biomarker in the rheumatoid arthritis classification criteria? *Clin Chem Lab Med*. 2019; 57(11):1668-1679.

[10] Farng E, Friedrich JB. Laboratory diagnosis of rheumatoid arthritis. *J Hand Surg Am*. 2011; 36(5):926-7.

[11] Wasserman AM. Diagnosis and management of rheumatoid arthritis. *Am Fam Physician*. 2011; 84(11):1245-52.

[12] Narváez J. Tratamiento de la artritis reumatoide [Treatment of rheumatoid arthritis]. *Med Clin (Barc)*. 2016; 147(4):176-80.

[13] Mulero Mendoza J. Tratamiento de la artritis reumatoide [Rheumatoid arthritis treatment]. *Rev Clin Esp*. 2004; 204(5):273-82.

[14] Solomon DH, Husni ME, Wolski KE, Wisniewski LM, Borer JS, Graham DY, Libby P, Lincoff AM, Lüscher TF, Menon V, Yeomans ND, Wang Q, Bao W, Berger MF, Nissen SE; PRECISION Trial Investigators. Differences in Safety of Nonsteroidal Antiinflammatory Drugs in Patients With Osteoarthritis and Patients With Rheumatoid Arthritis: A Randomized Clinical Trial. *Arthritis Rheumatol*. 2018; 70(4):537-546.

[15] Kapugi M, Cunningham K. Corticosteroids. *Orthop Nurs*. 2019; 38 (5):336-339.

[16] Buchman AL. Side effects of corticosteroid therapy. *J Clin Gastroenterol*. 2001; 33(4):289-294.

[17] Bath RK, Brar NK, Frouhar FA, Wu GY. A review of methotrexate-associated hepatotoxicity. *J Dig Dis*. 2014; 15(10):517-524.



[18] Naranjo Hernández A. Tratamiento de la artritis reumatoide [Treatment of rheumatoid arthritis]. *Med Clin (Barc)*. 2000; 114(20):777-83.

[19] Bullock J, Rizvi SAA, Saleh AM, Ahmed SS, Do DP, Ansari RA, Ahmed J. Rheumatoid Arthritis: A Brief Overview of the Treatment. *Med Princ Pract*. 2018;27(6):501-507.

[20] Scott LJ. Tocilizumab: A Review in Rheumatoid Arthritis. *Drugs*. 2017; 77(17):1865-1879.

[21] Wang L, Yang HJ, Tu R, Shen M, Liu D. In silico design of colchicine-based bioisosteric inhibitors of tubulin for the treatment of rheumatoid arthritis. *Mol Med Rep*. 2017; 16(4):4823-4828.

[22] Dhillon S. Tramadol/paracetamol fixed-dose combination: a review of its use in the management of moderate to severe pain. *Clin Drug Investig*. 2010; 30(10):711-38.

[23] Larson AM. Acetaminophen hepatotoxicity. *Clin Liver Dis*. 2007; 11(3):525-548.

[24] Boelsterli UA. Mechanisms of NSAID-induced hepatotoxicity: focus on nimesulide. *Drug Saf*. 2002; 25(9):633-648.

[25] Miner P Jr, Katz PO, Chen Y, Sostek M. Gastric acid control with esomeprazole, lansoprazole, omeprazole, pantoprazole, and rabeprazole: a five-way crossover study. *Am J Gastroenterol*. 2003; 98(12):2616-20.

[26] Trefts E, Gannon M, Wasserman DH. The liver. *Curr Biol*. 2017; 27(21):R1147-R1151.

[27] Radovanović-Dinić B, Tešić-Rajković S, Zivkovic V, Grgov S. Clinical connection between rheumatoid arthritis and liver damage. *Rheumatol Int*. 2018; 38(5):715-724.

[28] Garcia-Cortes M, Robles-Diaz M, Stephens C, Ortega-Alonso A, Lucena MI, Andrade RJ. Drug induced liver injury: an update. *Arch Toxicol.* 2020; 94(10):3381-3407.

[29] Genovese MC, Kremer JM, van Vollenhoven RF, Alten R, Scali JJ, Kelman A, Dimonaco S, Brockwell L. Transaminase Levels and Hepatic Events During Tocilizumab Treatment: Pooled Analysis of Long-Term Clinical Trial Safety Data in Rheumatoid Arthritis. *Arthritis Rheumatol.* 2017; 69(9):1751-1761.

[30] Agúndez JA, Lucena MI, Martínez C, Andrade RJ, Blanca M, Ayuso P, García-Martín E. Assessment of nonsteroidal anti-inflammatory drug-induced hepatotoxicity. *Expert Opin Drug Metab Toxicol.* 2011; 7(7):817-828.

[31] Rocco A, Compare D, Angrisani D, Sanduzzi Zamparelli M, Nardone G. Alcoholic disease: liver and beyond. *World J Gastroenterol.* 2014; 20(40):14652-14659.

[32] Howard DJ, Briggs LA, Pritsos CA. Oxidative DNA damage in mouse heart, liver, and lung tissue due to acute side-stream tobacco smoke exposure. *Arch Biochem Biophys.* 1998; 352(2):293-297.

[33] Daniel Faggella. Emerj [Internet]. [Consultado 5 Mar 2021]. Disponible en: <https://emerj.com/ai-sector-overviews/machine-learning-in-pharma-medicine/>.

[34] Sajda, Paul (2006). MACHINE LEARNING FOR DETECTION AND DIAGNOSIS OF DISEASE. *Annual Review of Biomedical Engineering*, 8(1), 537–565.

[35] MIT Clinical ML [Internet]. [Consultado 5 Mar 2021] Disponible en: <http://clinicalml.org/research/>.

[36] Marcus Roth. Emerj [Internet]. [Consultado 7 Mar 2021]. <https://emerj.com/ai-sector-overviews/machine-learning-radiology-current-applications/>.

[37] Gozalbes, R; JV, de Julián-Ortiz; Fito-López, C. Métodos computacionales en toxicología predictiva: aplicación a la reducción de ensayos con animales en el contexto de la legislación comunitaria REACH [Internet]. 2014 Consultado [Consultado 14 Mar



Diseño de software para obtener la interacción toxicológica entre fármacos en pacientes con enfermedades crónicas

2021]; 3. Disponible en: <https://www.redalyc.org/pdf/919/91932969008.pdf>

[38] SANOFI [Internet]. [Consultado 14 Mar 2021]. <https://campussanofi.es/e-professionals/noticias/que-es-real-world-evidence-y-big-data/>.

Lista de figuras

1.1 Riesgo de padecer artritis reumatoide por rango de edades en Colombia. http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0121-81232019000200083	16
1.2 Datos de prevalencia por todo el mundo de posibilidad de padecer artritis reumatoide. https://www.elsevier.es/es-revista-revista-colombiana-reumatologia-374-articulo-prevalencia-artritis-reumatoide-colombia-segun-S0121812319300167	17
3.2 Real Word Evidence. https://quantori.com/news/real-world-data-real-world-evidence	34
4.1 Datos de entrada reales para el entranamiento de los disintos tipos de red neuronal	42
4.2 Datos de salida reales de red neuronal simple	42
4.3 Arquitectura de red neuronal densa y no densa.	43
4.4 Código utilizado en el modelo de red neuronal simple.	43
4.5. Código utilizado en el entrenamiento de datos.	44
4.6. Red neuronal simple. Eje y: entrenamiento hasta estabilización, eje x: error.	45
4.7. Datos de prevalencia por todo el mundo de posibilidad de padecer artritis reumatoide. https://es.quora.com/C%C3%B3mo-aprende-una-red-neuronal-de-IA	46



4.8. Arquitectura con una capa oculta, eje y: entrenamiento hasta estabilización, eje x: error.	47
4.9. Modelo de red neuronal con capa oculta intermedia. https://artfromcode.wordpress.com/2017/04/18/red-neuronal-en-python-con-numpy-parte-1	48
4.10. Arquitectura con dos capas ocultas, eje y: entrenamiento hasta estabilización, eje x: error.	49
5.2. Iteraciones hasta estabilización de resultados. Eje.y: entrenamiento hasta estabilización, eje: error.....	52
5.3. Iteraciones hasta estabilización con una capa oculta. Eje.y: entrenamiento hasta estabilización, eje: error.....	53
5.4. Iteraciones hasta estabilización con con dos capas ocultas. Eje.y: entrenamiento hasta estabilización, eje: error.....	54
5.5. Resultados reales de salida de un paciente después de un año de tratamiento.....	55
5.6. Datos de salida del modelo de red neuronal con una capa oculta.....	56

5.7. Datos de salida del modelo de red neuronal con dos capas ocultas.....	56
5.8. Datos de salida del modelo de red neuronal simple	56

Lista de tablas

3.1.2. Fases de trabajo, tareas e instituciones responsables de llevarlas a cabo.....	32
3.2.2. Datos de entradas reales de pacientes.....	36
3.2.3. Datos reales de pacientes con sus respectivos medicamentos.....	36
3.2.4. Valores de analíticas.....	37
3.2.5. Valores de analíticas en % a predecir por nuestro programa.....	38
3.2.1. Especificaciones técnicas de la tarjeta gráfica.....	39
3.3.2. Especificaciones técnicas del procesador.....	39
5.2.1. Variables de entrada para el entrenamiento de datos.....	51
5.2.2. Valores toxicológicos de salida.....	51
5.3. Porcentaje de acierto de la predicción respecto a valor real de ejemplo.....	57