



Universidad
de La Laguna

Escuela Superior de
Ingeniería y Tecnología
Sección de Ingeniería Informática

Trabajo de Fin de Grado

Microsoft Azure Machine Learning para detección de distintos tipos de cáncer

Microsoft Azure Machine Learning for detecting cancers

Óscar Trujillo Acosta

La Laguna, 5 de julio de 2016

D. **Patricio García Báez**, con N.I.F. 43.356.987-D profesor contratado doctor adscrito al Departamento de Ingeniería Informática y de Sistemas de la Universidad de La Laguna, como tutor

D. **Carmen Paz Suárez Araujo**, con N.I.F. 43.640.373-N profesora Titular de Universidad adscrita al Departamento de Informática y Sistemas de la Universidad de Las Palmas de Gran Canarias, como co-tutora

C E R T I F I C A (N)

Que la presente memoria titulada:

“Microsoft Azure Machine Learning para detección de distintos tipos de cáncer.”

ha sido realizada bajo su dirección por D. **Óscar Trujillo Acosta**, con N.I.F. 42.222.212-P.

Y para que así conste, en cumplimiento de la legislación vigente y a los efectos oportunos firman la presente en La Laguna a 5 de julio de 2016.

Agradecimientos

En primer lugar quiero agradecer a mi tutor de trabajo, Patricio García por su implicación en éste y a mi cotutora Carmen Paz por interesarse por el tema de estudio y querer ayudar en todo lo posible.

También agradecer a mi compañero Alberto Fariña Barrera por haberme apoyado en aspectos puntuales del proyecto.

Por último y no menos importante, gracias a mis profesores, compañeros, amigos y familiares por todo el apoyo recibido durante mi etapa universitaria y por haberme ayudado a llegar aquí.

Licencia



© Esta obra está bajo una licencia de Creative Commons
Reconocimiento 4.0 Internacional.

Resumen

La computación tradicional es un campo que ha ido ligado a la historia de la especie humana desde sus inicios. Siglos y siglos de avance técnico han dado sus frutos y hoy en día disponemos de máquinas capaces de realizar cálculos computacionales muy superiores a lo que un ser humano es capaz de hacer. No obstante, no todo es una derrota y es que los animales seguimos poseyendo algo único: la capacidad de razonar. La naturaleza nos ha dotado de un sistema nervioso formado por una red neuronal muy compleja que nos permite aprender, razonar, deducir, etc.

En el último siglo el ser humano ha aunado esfuerzos en pos del desarrollo de sistemas que tratan de recrear esta forma de procesamiento humana, sistemas que han evolucionado hasta lo que hoy conocemos como: “Redes Neuronales Artificiales” o “RNA”.

Se propone con este proyecto el desarrollo de una red neuronal artificial aplicada a la medicina, en concreto su funcionalidad será la de detectar tres tipos de cáncer: Cáncer de mama, cáncer de piel y cáncer de pulmón.

Palabras clave: *Microsoft Azure Machine Learning, Redes Neuronales Artificiales, medicina, detección de cáncer, sistemas inteligentes.*

Abstract

The traditional computation is a science that has been linked to the history of the human species from its beginnings. Centuries of technical progress have paid off and today we have machines that can perform computations much higher than what a human being is able to do. However, not all is a beat because the animals still have something unique: the ability to reason. Nature has given us a nervous system consisting of a very complex neural network that allows us to learn, reason, deduce, etc.

In the last century, man has joined efforts for develop systems that try to recreate the human processing system. These systems have evolved to what we know today as Artificial Neural Networks or ANN.

This project propose development an intelligent system applied to medicine. Specific this system will be able to detect breast cancer, melanoma and cancer of the respiratory system.

Keywords: *Microsoft Azure Machine Learning, Artificial Neural Networks, medicine, cancer screening, intelligent systems.*

Índice General

Capítulo 1. Introducción	1
1.1 Antecedentes	1
1.2 Objetivos	1
1.3 Programa de Apoyo a Trabajos Finales Libres	1
1.4 Actividades a realizar	1
Capítulo 2. Redes Neuronales Artificiales, Datos y Herramientas	3
2.1 Redes Neuronales Artificiales	3
2.1.1 Historia	3
2.1.2 Funcionamiento	4
2.1.3 Estado actual	4
2.2 Base de Datos	5
2.2.1 Datos en RNA	5
2.2.2 Heuristic and Evolutionary Algorithms Laboratory	5
2.2.3 Estado de los datos	5
2.2.4 Resultados de obtenidos por HEAL	6
2.3 Herramientas	8
2.3.1 Herramientas de desarrollo de RNA	8
2.3.2 Microsoft Azure Machine Learning	9
Capítulo 3. Estudios realizados	16
3.1 Principios de los estudios	16
3.2 Resultados para cada tipo de cáncer analizado	19
3.2.1 Melanoma	19
3.2.2 Cáncer de pulmón	24
3.2.3 Cáncer de mama	29
3.3 Estudio de los sistemas bajo el subconjunto de test	34
3.3.1 Melanoma	35
3.3.2 Cáncer de pulmón	36
3.3.3 Cáncer de mama	37
3.4 Comparación de resultados con HEAL	38
3.4.1 Melanoma	38
3.4.2 Cáncer de pulmón	39
3.4.3 Cáncer de mama	40

Capítulo 4. Conclusiones y líneas futuras	41
Capítulo 5. Summary and Conclusions	42
Capítulo 6. Presupuesto	43
6.1 Conceptos	43
Bibliografía	44

Índice de figuras

Figura 2.1. Resultados HEAL cáncer de mama	6
Figura 2.2. Resultados HEAL melanoma.....	7
Figura 2.3. Resultados HEAL cáncer de pulmón.....	7
Figura 2.4. Clasificación de Algoritmos de Azure.....	9
Figura 2.5. Árbol de decisión	11
Figura 2.6. Grafo acíclico dirigido	11
Figura 2.7. Grafo uso de Azure ML	12
Figura 2.8. Ejemplo resultados Azure.....	12
Figura 2.9. Ejemplo gráfica curva ROC Azure.....	13
Figura 2.10. Azure ML módulo evaluador.....	14
Figura 2.11. Ejemplo gráfica P/R Azure	14
Figura 3.1. Esquema de las pruebas.	16
Figura 3.2. Gráfico bloque de prueba en Azure ML.	17
Figura 3.3. Gráficas de los resultados del bloque 4 para el melanoma. (a) Curva ROC. (b) Curva P/R.....	19
Figura 3.4. Gráficas de los resultados del bloque 4 para el melanoma. (a) Curva ROC. (b) Curva P/R.....	20
Figura 3.5. Gráficas de los resultados del bloque 3 para el melanoma. (a) Curva ROC. (b) Curva P/R.....	21
Figura 3.6. Gráficas de los resultados del bloque 4 para el melanoma. (a) Curva ROC. (b) Curva P/R.....	22
Figura 3.7. Gráficas de los resultados del bloque 1 para el cáncer de pulmón. (a) Curva ROC. (b) Curva P/R.	24
Figura 3.8. Gráficas de los resultados del bloque 2 para el cáncer de pulmón. (a) Curva ROC. (b) Curva P/R.	25
Figura 3.9. Gráficas de los resultados del bloque 3 para el cáncer de pulmón. (a) Curva ROC. (b) Curva P/R.	26
Figura 3.10. Gráficas de los resultados del bloque 4 para el cáncer de pulmón. (a) Curva ROC. (b) Curva P/R.	27
Figura 3.11. Gráficas de los resultados del bloque 1 para el cáncer de mama. (a) Curva ROC. (b) Curva P/R.	29
Figura 3.12. Gráficas de los resultados del bloque 2 para el cáncer de mama. (a) Curva ROC. (b) Curva P/R.	30

Figura 3.13. Gráficas de los resultados del bloque 3 para el cáncer de mama. (a) Curva ROC. (b) Curva P/R.	31
Figura 3.14. Gráficas de los resultados del bloque 4 para el cáncer de mama. (a) Curva ROC. (b) Curva P/R.	32
Figura 3.15. Esquema de módulos subconjunto de test.	34
Figura 3.16. Gráficas del mejor modelo para el conjunto de test de los datos de melanoma. (a) Curva ROC. (b) Curva P/R.	35
Figura 3.17. Gráficas del mejor modelo para el conjunto de test de los datos de cáncer de pulmón. (a) Curva ROC. (b) Curva P/R.....	36
Tabla 3.18. Resultados del mejor modelo para el conjunto de test de los datos de cáncer de pulmón.	36
Figura 3.19. Gráficas del mejor modelo para el conjunto de test de los datos de cáncer de pulmón. (a) Curva ROC. (b) Curva P/R.....	37

Índice de tablas

Tabla 2. 1. Configuración redes melanoma bloque 1.....	6
Tabla 2. 2. Configuración redes melanoma bloque 1.....	6
Tabla 3.1. Configuración de los modelos del bloque 4 para el melanoma	19
Tabla 3.2. Resultados de los modelos del bloque 4 para el melanoma.....	19
Tabla 3.3. Configuración de los modelos del bloque 4 para el melanoma	20
Tabla 3.4. Resultados de los modelos del bloque 4 para el melanoma.....	20
Tabla 3.5. Configuración de los modelos del bloque 4 para el melanoma	21
Tabla 3.6. Resultados de los modelos del bloque 3 para el melanoma.....	21
Tabla 3.7. Configuración de los modelos del bloque 4 para el melanoma	22
Tabla 3.8. Resultados de los modelos del bloque 4 para el melanoma.....	22
Tabla 3.9. Comparación final de modelos sobre el conjunto de entrenamiento y validación del melanoma.....	23
Tabla 3.10. Configuración de los modelos del bloque 1 para el cáncer de pulmón ...	24
Tabla 3.11. Resultados de los modelos del bloque 1 para el cáncer de pulmón	24
Tabla 3.12. Configuración de los modelos del bloque 2 para el cáncer de pulmón ...	25
Tabla 3.13. Resultados de los modelos del bloque 2 para el cáncer de pulmón	25
Tabla 3.14. Configuración de los modelos del bloque 3 para el cáncer de pulmón ...	26
Tabla 3.15. Resultados de los modelos del bloque 3 para el cáncer de pulmón	26
Tabla 3.16. Configuración de los modelos del bloque 4 para el cáncer de pulmón ...	27
Tabla 3.17. Resultados de los modelos del bloque 4 para el cáncer de pulmón	27
Tabla 3.18. Comparación final de modelos sobre el conjunto de entrenamiento y validación del cáncer de pulmón	28
Tabla 3.19. Configuración de los modelos del bloque 1 para el cáncer de mama.....	29
Tabla 3.20. Resultados de los modelos del bloque 1 para el cáncer de mama	29
Tabla 3.21. Configuración de los modelos del bloque 2 para el cáncer de mama.....	30
Tabla 3.22. Resultados de los modelos del bloque 2 para el cáncer de mama	30
Tabla 3.23. Configuración de los modelos del bloque 3 para el cáncer de mama.....	31
Tabla 3.24. Resultados de los modelos del bloque 3 para el cáncer de mama	31
Tabla 3.25. Configuración de los modelos del bloque 4 para el cáncer de mama.....	32
Tabla 3.26. Resultados de los modelos del bloque 4 para el cáncer de mama	32

Tabla 3.27. Comparación final de modelos sobre el conjunto de entrenamiento y validación del cáncer de mama	33
Tabla 3.28. Resultados del mejor modelo para el conjunto de test de los datos de melanoma	35
Tabla 3.29. Resultados del mejor modelo para el conjunto de test de los datos de cáncer de pulmón	36
Figura 3.19. Gráficas del mejor modelo para el conjunto de test de los datos de cáncer de pulmón (a) Curva ROC (b) Curva P/R.....	37
Tabla 3.30. Resultados del mejor modelo para el conjunto de test de los datos de cáncer de mama.....	37
Tabla 3.31. Comparación melanoma con resultados de HEAL, subconjunto de entrenamiento y validación	38
Tabla 3.32. Comparación melanoma con resultados de HEAL, subconjunto de tests	38
Tabla 3.33. Comparación cáncer de pulmón con resultados de HEAL, subconjunto de entrenamiento y validación	39
Tabla 3.34. Comparación cáncer de pulmón con resultados de HEAL, subconjunto de tests	39
Tabla 3.35. Comparación cáncer de mama con resultados de HEAL, subconjunto de entrenamiento y validación	40
Tabla 3.36. Comparación cáncer de mama con resultados de HEAL, subconjunto de tests	40
Tabla 6. 1. Presupuesto.....	43

Capítulo 1.

Introducción

1.1 Antecedentes

Las redes neuronales artificiales han experimentado un importante auge en los últimos años, grandes empresas como Microsoft, Amazon, IBM o Google han apostado por desarrollar plataformas relacionadas con este campo, sistemas, por lo general, especialmente diseñados para compartir la información generada y utilizarla como piedra angular en su propio desarrollo.

En el campo de la medicina el estudio de redes neuronales artificiales no es nada nuevo, existen multitud de desarrollos que tratan de facilitar nuevos sistemas de detección precoz de enfermedades para optimizar y facilitar la labor de los profesionales del sector.

1.2 Objetivos

El objetivo de este trabajo de fin de grado es realizar un estudio sobre el uso de sistemas inteligentes en la predicción de cáncer con determinada herramienta. Se analizarán principalmente diferentes tipos de redes neuronales artificiales y se compararán con otros modelos utilizados en la inteligencia artificial como son los árboles de decisión en búsqueda del mejor modelo de detección de cáncer con el conjunto de datos que se tiene.

1.3 Programa de Apoyo a Trabajos Finales Libres

Este trabajo de fin de grado está adherido al programa de apoyo a trabajos finales libres de la Oficina de Software Libre de La Universidad [1].

1.4 Actividades a realizar

Se establece una serie de tareas que servirán para identificar el alcance de este trabajo de fin de grado:

- a) Preparación
 - a. Búsqueda de información sobre redes neuronales
 - b. Estudio sobre software relacionado con redes neuronales
 - c. Estudio del trabajo del laboratorio HEAL
- b) Preprocesado de datos
 - a. Análisis de los datos
 - b. Elección de herramienta de preprocesado
 - c. Realizar preprocesado
 - d. Seleccionar particiones del conjunto de datos
- c) Diseño y desarrollo del diagnosticador
 - a. Elegir modelos
 - b. Diseño del sistema
 - c. Entrenamiento y validación
 - d. Optimización de los parámetros del diagnosticador
- d) Evaluación del diagnosticador
 - a. Obtener porcentaje de acierto real
 - b. Crear gráficas de errores y precisión
 - c. Comparar con resultado del equipo austriaco
- e) Finalización
 - a. Preparación de la memoria del TFG
 - b. Preparación de la defensa de la memoria

Capítulo 2.

Redes Neuronales Artificiales, Datos y Herramientas

2.1 Redes Neuronales Artificiales

En este capítulo se realiza una breve introducción a las redes neuronales artificiales, se explica cuáles son los datos que se usan en los estudios y se muestran diferentes herramientas informáticas para el manejo de redes neuronales artificiales.

2.1.1 Historia

La inteligencia artificial [2] o IA es un área multidisciplinaria estrechamente relacionada con la ingeniería informática que estudia métodos para resolver problemas utilizando la forma de procesamiento humana como paradigma. Este objetivo básico de la IA nace en la civilización Griega cuando Aristóteles trata de recoger una serie de reglas que describen el comportamiento del procesamiento humano. No obstante, el termino Inteligencia Artificial es acuñado en 1956 durante la conferencia de Dartmouth.

La IA se sirve de la capacidad de cómputo de las CPU actuales para procesar información y gracias a esto consigue resultados altamente veraces y veloces en determinados problemas. No obstante existen otro tipo de problemas en los que un sistema inteligente no puede alcanzar un resultado óptimo, si pudiendo hacerlo un cerebro humano. Esto se debe a que el procesamiento en el cerebro es fruto de la autoorganización de millones de procesadores elementales llamados neuronas y no a la potencia secuencial de cada una de ellas como plantea la computación tradicional.

De esta manera surge una ruptura con el funcionalismo de la estructura de cómputo clásica dando origen a las redes neuronales artificiales [3][4] cuya idea de partida es que para ejecutar aquel tipo de tareas que más eficazmente resuelve un cerebro, puede resultar interesante copiar su estructura hardware.

Una red neuronal artificial, también llamada RNA por sus iniciales, es un paradigma informático de procesamiento de información inspirado en los sistemas neuronales biológicos.

2.1.2 Funcionamiento

Las RNA tratan la información de forma similar a como lo hace un cerebro humano, aprenden con ejemplos y no son programadas para que hagan una tarea específica.

El aprendizaje en un sistema biológico se realiza mediante el ajuste de las conexiones sinápticas entre las neuronas, algo que también se hace en las redes neuronales artificiales. De esta manera una RNA es capaz de reconocer patrones y aprender de sus errores.

2.1.3 Estado actual

Durante los últimos años muchas son las grandes empresas del campo de la informática como Google [5], IBM [6], Microsoft [7] o Twitter [8] que están apostando por el desarrollo de sistemas basados en redes neuronales. La cantidad de posibilidades que ofrece esta tecnología y lo poco explorada que se encuentra son motivantes muy grandes para su estudio y puesta en práctica.

Son muchas las utilidades que puede ofrecer un sistema basado en redes neuronales, concretamente en la medicina [9] se han desarrollado diferentes aplicaciones con fines tales como la ayuda al doctor en el diagnóstico de enfermedades o el asesoramiento directo a pacientes según sus síntomas [10]. En ambas situaciones se trata de un sistema que puede resultar de gran utilidad y su buen uso puede significar un avance significativo en este campo. No obstante, se tiene que tener en cuenta en todo momento que estos sistemas no se pueden considerar, aún, cien por cien efectivos.

2.2 Base de Datos

2.2.1 Datos en RNA

La propia naturaleza de funcionamiento de las redes neuronales artificiales requiere una gran cantidad de datos de entrenamiento para ser ajustada y que consiga funcionar correctamente. Los datos necesarios para este estudio son de carácter sensible y la disponibilidad certera de los mismos no es alta ya que se tratan de datos clínicos personales.

Esta dificultad se ha visto salvada gracias al contacto establecido con el “Heuristic and Evolutionary Algorithms Laboratory”, también llamado HEAL [11].

2.2.2 Heuristic and Evolutionary Algorithms Laboratory

El Heuristic and Evolutionary Algorithms Laboratory (HEAL) es un grupo de investigadores austriaco que ha realizado un estudio similar en cuanto a objetivo al pretendido en este trabajo, han desarrollado su propio sistema de detección de cáncer usando optimización genética de árboles de decisión. Los árboles de decisión no son más que otro paradigma de predicción usado en el ámbito de la inteligencia artificial. Este grupo de investigación nos ha cedido los datos usados en este estudio para poder entrenar a nuestros sistemas y a la vez tenemos los resultados en porcentaje de acierto obtenidos por ellos, algo de gran utilidad si queremos comparar nuestros resultados finales con los obtenidos en pruebas de naturaleza similar.

2.2.3 Estado de los datos

Los estudios realizados se han hecho a partir de los ficheros .csv que ha facilitado el HEAL para el desarrollo de este proyecto, los cuales contienen los datos de un total de casi 4000 pacientes y 30 variables para cada uno de ellos procedentes del análisis sanguíneo del sujeto. Estas variables son:

- Cáncer (1/0), Edad, sexo, AST, ALT, GT37, BUN, Creatinina, CH37, LD37, Ácido Úrico, Bilirrubina, Colesterol, Colesterol HDL, Hierro, Ferritina, Transferrina, CRP, BSG1, Leucocitos, Neutrófilos, Linfocitos, Monocitos, Eosinófilos, Basófilos, Eritrocitos, Hemoglobina, Hematocritos, MCV y Plaquetas.

El conjunto de estos datos se encuentra dividido en tres ficheros, uno por cada tipo de cáncer con 706 entradas para el fichero de cáncer de mama, 905 entradas para el fichero de melanoma y 2363 entradas para el fichero de cáncer de pulmón. Cada fila del fichero es una entrada que representa a un paciente.

No obstante para el buen funcionamiento del sistema se ha decidido eliminar ciertas columnas de información donde la mayor parte de los datos eran nulos o estaban corruptos así como ciertas filas de datos por el mismo motivo. El resto de datos desconocidos o nulos se han resuelto con la media de los datos de su columna y en ningún caso se han normalizado las entradas.

En la tabla 2.1 se indica el ratio de pacientes sanos y enfermos para cada tipo de cáncer analizado y en la tabla 2.2 se muestra el porcentaje de filas descartadas para el análisis realizado para este trabajo de fin de grado por cada tipo de cáncer.

	Cáncer de mama	Melanoma	Cáncer de pulmón
Pacientes enfermos	54,11%	39,30%	31,48%
Pacientes sanos	45,89%	60,70%	68,52%
Total	697	799	1995

Tabla 1. 2. Configuración redes melanoma bloque 1

Tipo de cáncer	Porcentaje de filas descartadas
Melanoma	11.71%
Cáncer de pulmón	15.57%
Cáncer de mama	1.27%

Tabla 3. 4. Configuración redes melanoma bloque 1

2.2.4 Resultados de obtenidos por HEAL

Además de los datos de análisis sanguíneos, HEAL nos ha cedido sus resultados, frutos de una amplia investigación en el tema. Estos resultados resultan de gran utilidad para este estudio ya que nos sirve para comparar los resultados obtenidos con un estudio similar hecho con los mismos datos.

Criteria	Training			Test		
	Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity
Best Accuracy	81.87%	89.79%	72.53%	74.36%	84.29%	62.65%
Best Sensitivity	76.63%	95.29%	54.63%	74.08%	92.67%	52.16%
Best of Both	77.05%	95.55%	55.25%	74.36%	91.88%	53.70%

Figura 1.5. Resultados HEAL cáncer de mama

	Training			Test		
Criteria	Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity
Best Accuracy	83.20%	79.52%	86.39%	73.92%	66.19%	80.62%
Best Sensitivity	83.65%	81.43%	85.57%	73.04%	67.14%	78.14%
Best of Both	83.20%	80.24%	85.77%	73.15%	66.67%	78.76%

Figura 2.6. Resultados HEAL melanoma

	Training			Test		
Criteria	Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity
Best Accuracy	92.47%	87.25%	96.27%	87.26%	76.71%	94.95%
Best Sensitivity	92.85%	89.06%	95.61%	87.64%	78.31%	94.44%
Best of Both	92.93%	88.96%	95.83%	87.30%	77.61%	94.37%

Figura 3.7. Resultados HEAL cáncer de pulmón

En las figuras 2.1, 2.2 y 2.3 se muestran los resultados finales obtenidos por el HEAL y seguidamente se encuentran la relación de índices que han usado para mostrar sus resultados:

$$\text{Sensitivity} = \text{TP} / (\text{TP} + \text{FN})$$

$$\text{Specifity} = \text{TN} / (\text{TN} + \text{FP})$$

$$\text{Accuracy} = (\text{TP} + \text{TN}) / (\text{TP} + \text{TN} + \text{FP} + \text{FN})$$

Donde TP son los verdaderos positivos, FP los falsos positivos, FN los falsos negativos y TN los verdaderos negativos

2.3 Herramientas

2.3.1 Herramientas de desarrollo de RNA

Existen multitud de herramientas dedicadas a la implementación de RNAs, para la realización de este trabajo se ha hecho un amplio estudio del campo y se han seleccionado un total de 7 herramientas que se han dividido en dos grupos. Por un lado librerías, generalmente de código abierto, que ofrecen un amplio abanico de algoritmos y funcionalidades en distintos lenguajes de programación entre los que destacan Python, C/C++ y R. Por otro lado se encuentran una serie de entornos de desarrollo ya sean en versión software o web que ofrecen una serie de algoritmos de aprendizaje genéticos junto con una GUI diseñada para facilitar cada paso.

a) Librerías:

- NeuroLab: Simple y potente librería para RNAs bajo Python. Incluye modelos supervisados, no supervisados y recurrentes [12].
- NeuralNet: Se trata de una librería en lenguaje R que incluye diferentes algoritmos de aprendizaje supervisado [13].
- PyBrain: Al igual que la anterior se trata de una librería Python que incluye modelos supervisados, no supervisados y recurrentes. Tiene la ventaja de que separa la creación de la red neuronal del hecho de entrenarla, cosa que es imposible en NeuroLab. Además PyBrain posee su propio tipo de datos [14].
- Fast Artificial Neural Network: También conocida como FANN es una librería de código abierto escrita en C reconocida por su alta velocidad de procesamiento [15].
- Tensor Flow: Librería de aprendizaje profundo OpenSource desarrollada por Google usada por la compañía en su famoso buscador y en su asistente personal Google Now [16].

b) Entornos de desarrollo:

- IBM SPSS Modeler: Se trata de un paquete software dedicado al modelado de datos que también proporciona diferentes algoritmos de aprendizaje como RNAs [17].
- Weka: Software de escritorio para el aprendizaje automático y minería de datos diseñado en Java. Herramienta versátil que soporta multitud de tareas de procesamiento de datos y un amplio conjunto de algoritmos de aprendizaje de máquina, incluidos varios de RNAs. Incorpora la posibilidad de ser utilizado mediante llamadas a su librería [18].

- **Microsoft Azure Machine Learning:** Microsoft plantea esta herramienta como un servicio de cloud computing. Accedemos a nuestro espacio de trabajo a través de un portal web donde tenemos alojadas las diferentes redes neuronales artificiales así como los datos que usamos. Esta característica nos ofrece la posibilidad de ejecutar los costos algoritmos de aprendizaje en servidores de la compañía especialmente diseñados para esto. Además, el abanico de modelos de redes neuronales que ofrece es muy amplio y la versión gratuita de uso no es nada limitada. Asimismo destacar también las múltiples funcionalidades extra que incorpora como la posibilidad de usar nuestros propios scripts en R o Python y un potente conjunto de herramientas de modelado de datos. Por último añadir también que la interfaz que ofrece es muy cómoda e intuitiva [19].

2.3.2 Microsoft Azure Machine Learning

Microsoft Azure Machine Learning es una herramienta que sigue en desarrollo, no obstante ya cuenta con un abanico amplio de algoritmos de aprendizaje que han clasificado en los 5 tipos que pueden ver en la figura 2.4.

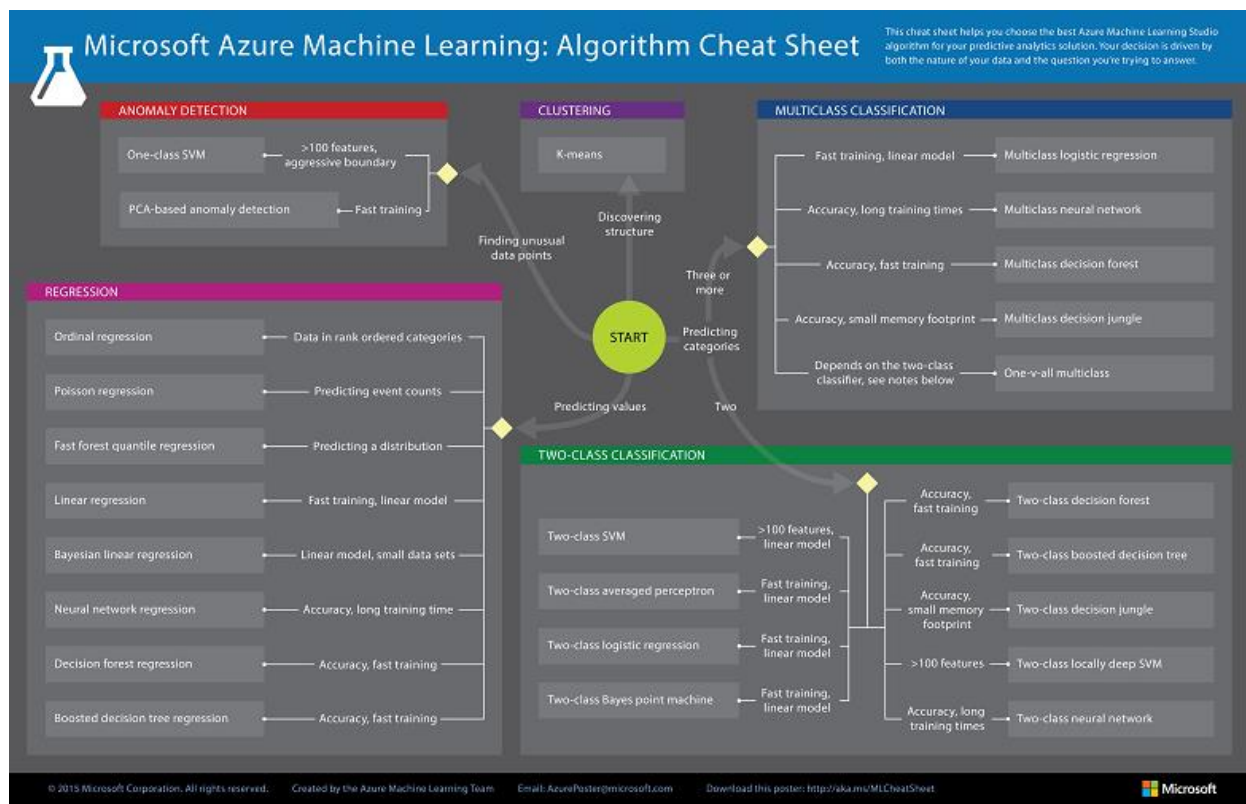


Figura 4.8. Clasificación de Algoritmos de Azure

Sin embargo, se analizarán tan solo los algoritmos supervisados de tipo “two-class classification” ya que es la categoría que se adapta a las necesidades del estudio, es decir, poder predecir si un sujeto tiene cáncer o no.

Modelos de redes neuronales de Azure ML:

- Two-Class Averaged Perceptron: Se trata del algoritmo de clasificación más simple, el perceptrón. Se trata una versión temprana de lo que serían las redes neuronales. Las entradas se clasifican en varias salidas posibles en base a una función lineal y después de ajustar una serie de pesos se obtiene una salida [20].
- Two-Class SVM: SVM es el acrónimo de Support Vector Machine o máquina de vector de soporte en español. Dado un conjunto de puntos en el espacio, los cuales cada uno de ellos pertenece a una de dos posibles categorías, un algoritmo basado en SVM construye un modelo capaz de predecir si un punto nuevo (cuya categoría desconocemos) pertenece a una categoría o a la otra [21].
- Two-Class Neural Network: Modelo de red neuronal clásica. Este algoritmo permite definir el número de capas de la red neuronal, número de neuronas por capa y como están definidas sus relaciones [22].
- Two-Class Locally Deep SVM: Los algoritmos de aprendizaje profundo contrastan con los algoritmos de aprendizaje poco profundo por el número de transformaciones aplicadas a la señal mientras se propaga desde la capa de entrada a la capa de salida. Cada una de estas transformaciones incluyen parámetros que se pueden entrenar como pesos y umbrales. Esta es la fórmula de optimización que usa este algoritmo [23]:

$$\min_{W, \theta, \theta'} P(W, \theta, \theta') = \frac{\lambda_W}{2} Tr(W^t W) + \frac{\lambda_\theta}{2} Tr(\theta^t \theta) + \frac{\lambda_{\theta'}}{2} Tr(\theta'^t \theta') + \sum_{i=1}^N L(y_i, \phi_L^t(x_i) W^t x_i)$$

Algoritmos basados en árboles de decisión

Los árboles de decisión [24] son construcciones lógicas que sirven para categorizar y representar condiciones para la resolución de un problema.

Para completar este estudio resulta de utilidad hacer una comparación de los resultados obtenidos de las redes neuronales artificiales con resultados obtenidos de otro modelo inteligente, como son los árboles de decisión.

- Two-Class Boosted Decision Tree: Se trata de un algoritmo supervisado que es capaz de crear un número determinado de árboles que conectados secuencialmente son capaces de corregir los errores del inmediatamente anterior. De esta manera se toma una decisión de clasificación lo más acertada posible. En la figura 2.5. se explica gráficamente cómo se comporta in árbol de este tipo.

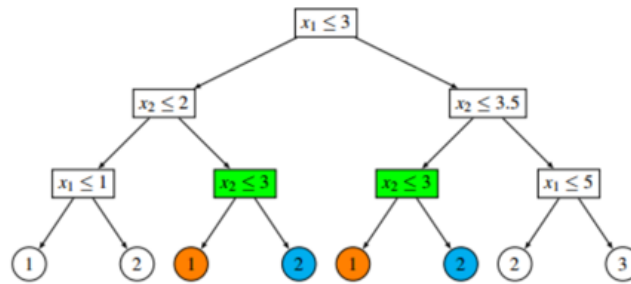


Figura 5.9. Árbol de decisión

- Two-Class Decision Forest: Este algoritmo crea múltiples árboles de decisión y después vota por la salida más “popular” entre las salidas de los diferentes árboles.
- Two-Class Decision Jungle: Se basa en la idea del algoritmo anterior pero en vez de usar árboles de decisión como unidades de voto usa un grafo acíclico dirigido (DAG por sus iniciales en inglés). El DAG es un sistema que trata de minimizar la gran cantidad de memoria generado por los árboles de decisión. En la figura 2.6 se observa cómo se comporta un sistema de este tipo.

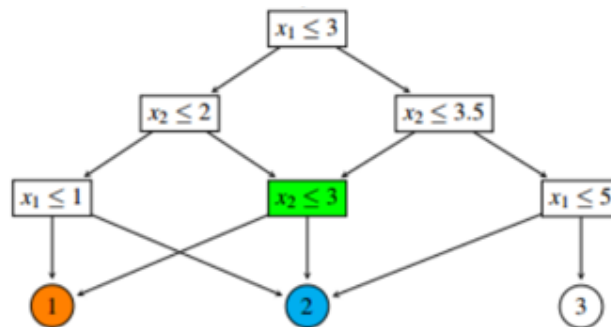


Figura 6.10. Grafo acíclico dirigido

Funcionamiento de la herramienta

Microsoft ofrece una intuitiva interfaz de módulos que a través de un sistema de “drag and drop” se podrá usar, ajustar e interconectar. En la figura 2.7 se puede observar un diagrama de trabajo básico de la herramienta.

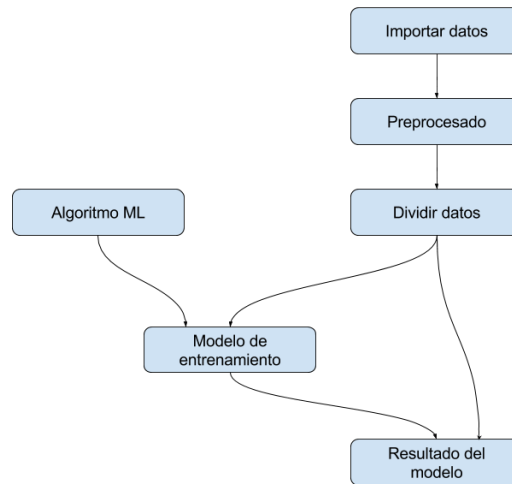


Figura 7.11. Grafo uso de Azure ML

Como se puede observar en la figura 2.7, la herramienta comienza importando los datos, preprocesándolos y dividiéndolos en dos subconjuntos para el posterior entrenamiento del modelo. Por otro lado se ha de elegir cuál es el algoritmo que utilizaremos para nuestro proyecto y finalmente obtendremos un resultado.

A estas funcionalidades básicas se le pueden sumar un amplio abanico de posibilidades como pueden ser la representación de los resultados en gráficas o la utilización de lenguaje Python o R en cualquier parte del proceso de modelado de datos.

Análisis de los gráficos y datos de Azure

En la figura 2.8 se muestra un ejemplo de cómo Azure ML muestra las estadísticas obtenidas de un modelo por defecto.

True Positive	False Negative	Accuracy	Precision	Threshold	AUC
46	51	0.650	0.622	0.5	0.683
False Positive	True Negative	Recall	F1 Score		
28	101	0.474	0.538		
Positive Label	Negative Label				
1	0				

Figura 8.12. Ejemplo resultados Azure

Los siguientes son los índices usados por Azure ML para mostrar los resultados obtenidos en las pruebas:

- **Accuracy:**
Es simplemente el ratio de predicciones correctas.
 $TP / (TP + FP + TN + FN)$

- Precision o precisión:
Porcentaje de predicciones positivas correctas:
 $TP / (TP + FP)$
- Recall o sensibilidad:
Es la relación entre las predicciones positivas correctas y los falsos negativos.
 $TP / (TP + FN)$
- F1 Score:
Es la media ponderada de la precisión y la sensibilidad.
 $2 * (Recall * Precision) / Recall + Precision$
- Threshold o umbral:
El modelo de entrenamiento genera un número real entre 0 y 1 que después es usado como threshold o umbral para determinar si dicho número se traduce en un “positive” o en un “negative”. Por defecto este valor se marca como 0,5 lo que significa que todos los índices recogidos mayor que este valor serán interpretados como positivos y los menores como negativos. No obstante Azure nos permite la posibilidad de modificarlo dinámicamente.
- AUC:
AUC es un valor entre 0 y 1 que representa el área debajo de la curva ROC. Cuanto mayor sea significa, por lo general, que el modelo tiene un porcentaje de acierto mayor.

Curva ROC:

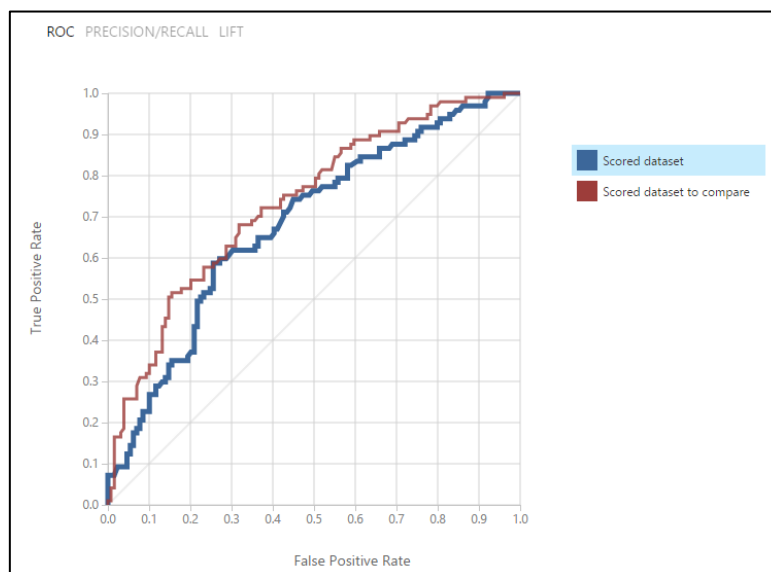


Figura 9.13. Ejemplo gráfica curva ROC Azure

En gráfica de la figura 2.9 dos modelos de entrenamientos son comparados, la curva azul representa al modelo conectado al módulo evaluador de Azure ML por la izquierda del mismo y la roja por la derecha como se muestra en la figura 2.10. Cada posición en la línea representa el resultado del modelo para cada threshold o umbral seleccionado. La esquina inferior izquierda (0,0) representa un umbral de 1 mientras que la esquina superior derecha (1,1) representa un umbral de 0. Las dimensiones donde se mueve son un eje vertical que representa los TP y un eje horizontal que representa los FP.

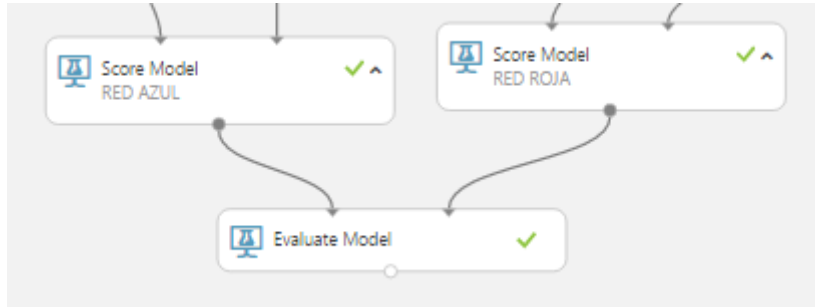


Figura 10.14. Azure ML módulo evaluador

Al cambiar el umbral, disminuye la frecuencia de un tipo de error a expensas de aumentar el otro tipo. A la hora de comparar dos modelos nos interesa saber cuál curva se acerca más a la esquina superior izquierda ya que este modelo proporciona una mejor predicción.

Precision/recall:

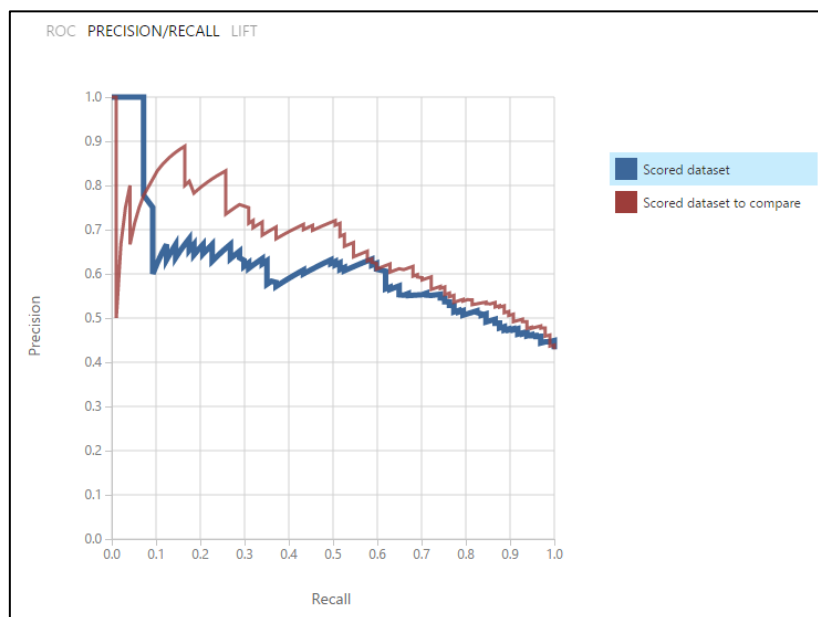


Figura 11.15. Ejemplo gráfica P/R Azure

En la gráfica de la figura 2.11 ambas curvas se mueven bajo los ejes vertical “precisión” y horizontal “recall”. Cada punto representa el porcentaje de acierto para datos positivos y negativos cuando variamos el umbral. La curva que más se acerque a la esquina superior derecha será la que ostente mejores resultados ya que se traduce en el modelo con más aciertos.

Capítulo 3. Estudios realizados

3.1 Principios de los estudios

Los estudios se han realizado independientemente para cada tipo de cáncer, no obstante todos siguen un patrón, el cual será explicado en esta sección.

En la figura 3.1 se muestra de forma gráfica cómo se han desarrollado las pruebas con la interfaz de Azure ML. Este sistema es análogo para cada tipo de cáncer y tan solo cambian los datos introducidos y los resultados obtenidos.

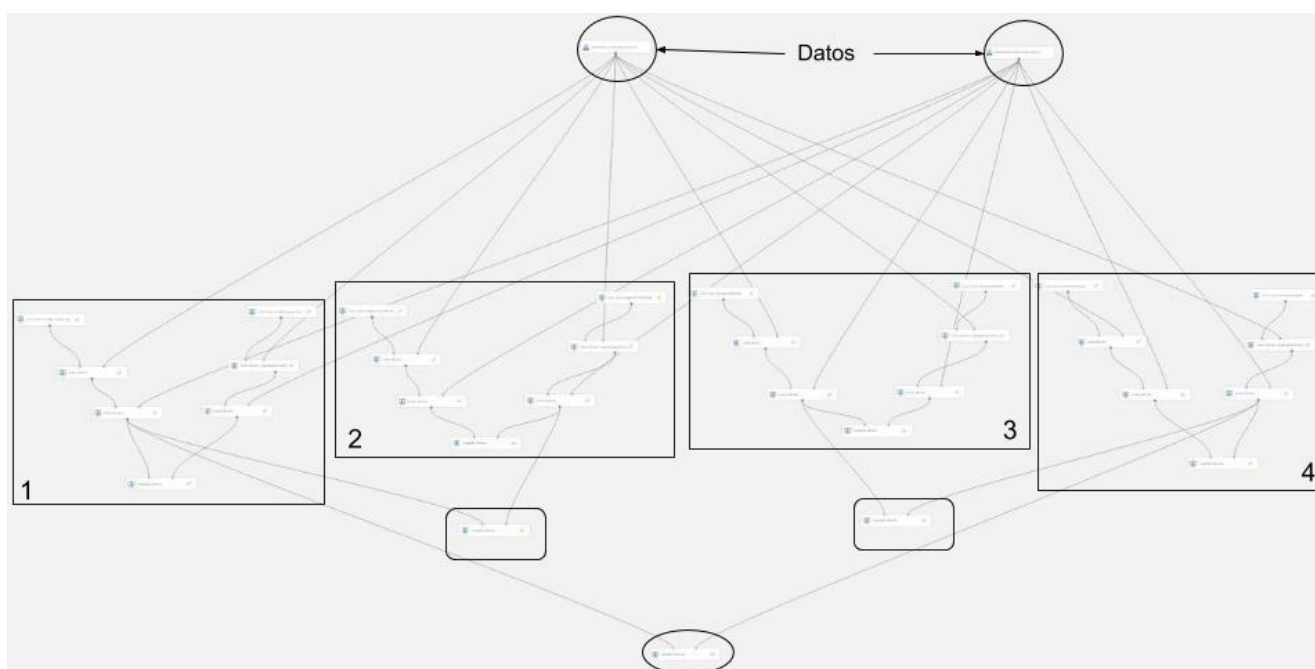


Figura 1.1. Esquema de las pruebas

Los dos módulos superiores son las entradas de datos. Representan al subconjunto de datos de entrenamiento (izquierda) y de validación (derecha).

En los siguientes cuatro grandes bloques de módulos se encuentran los modelos inteligentes definidos. Cada bloque, a su vez, tiene una estructura similar al resto, expuesta gráficamente en la figura 3.2.

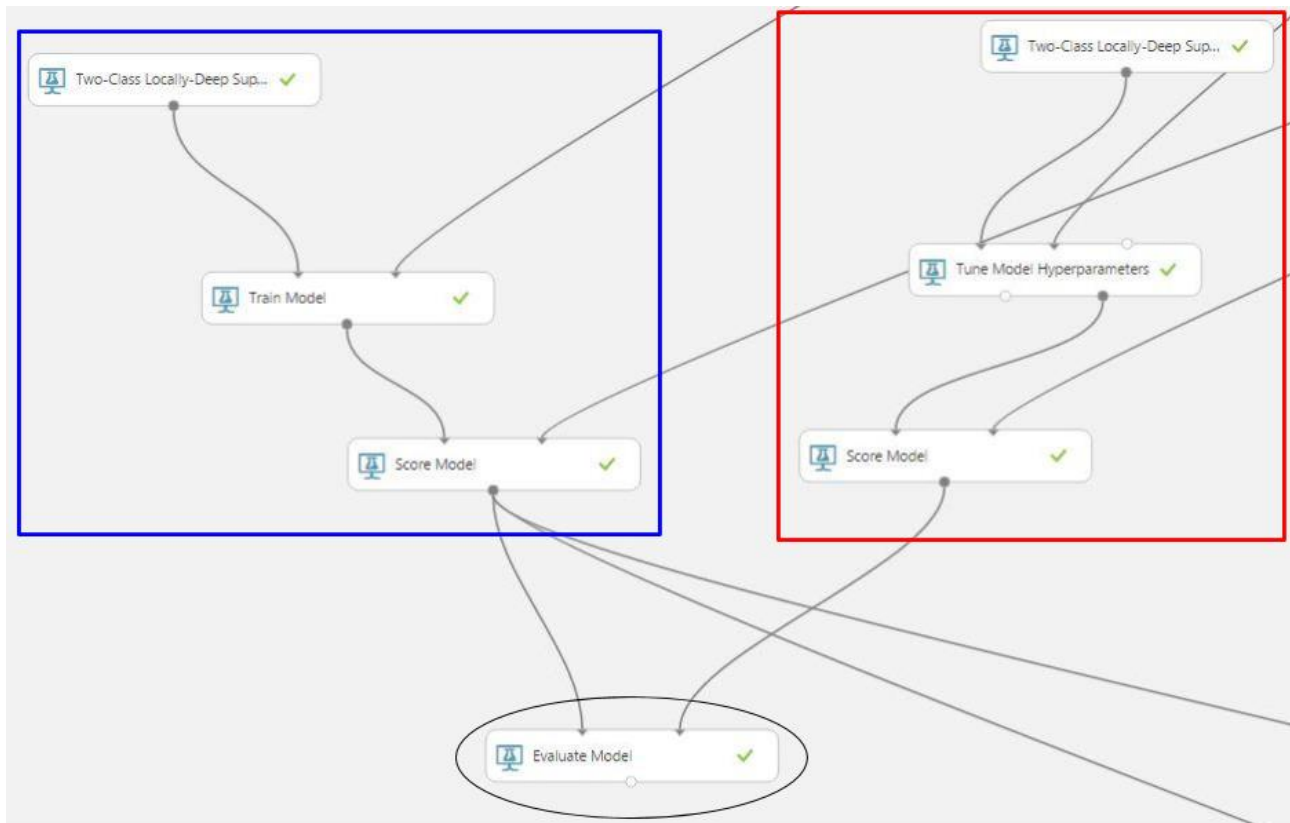


Figura 2.2. Gráfico bloque de prueba en Azure ML

En cada bloque se sigue una estructura similar a la mostrada en la Figura 3.2, con la diferencia de que en cada bloque se cambia el modelo de red neuronal definida, en este ejemplo es un modelo “Two-Class Locally-Deep Support Vector Machine”.

Como pueden observar se han dividido en dos cada bloque, esto se debe a que dentro de cada bloque se han creado dos redes neuronales distintas que después son comparadas en el módulo que se observa debajo, marcado con una elipse. Los colores definidos ayudarán después a la hora de comparar los resultados de tal forma que siempre la red azul será la de la izquierda y la roja la de la derecha.

La red azul en cada caso es el modelo definido con los parámetros por defecto de Azure ML y la red roja es una red obtenida de la comparación de veinte modelos definidos con parámetros aleatorios en búsqueda del mayor porcentaje de aciertos. Una vez entrenadas estas dos redes se comparan en el módulo “Evaluate Model”.

Esta primera comparación entre modelos se realiza dentro de cada bloque para obtener el mejor modelo de red. Como se ha mencionado anteriormente en cada bloque se define un modelo:

- Bloque 1: Two-Class Locally-Deep SVM.
- Bloque 2: Two-Class Support Vector Machine.
- Bloque 3: Two-Class Neural Network.
- Bloque 4: Two-Class Decision Jungle.

Una vez obtenida la configuración de modelo más óptimo para cada tipo de modelo se realiza una comparación entre sí para obtener qué modelo es el que obtiene los mejores resultados.

Para decidir qué RNA obtiene mejores resultados se analiza el global de índices obtenidos en las pruebas, no obstante, nos centraremos en los valores de Accuracy, Recall y Precision.

3.2 Resultados para cada tipo de cáncer analizado

En esta sección se mostrarán los resultados obtenidos en las pruebas realizadas para cada tipo de cáncer.

3.2.1 Melanoma

Para el análisis del bloque 1 correspondiente a los datos de melanoma con el modelo “Two-Class Locally-Deep Support Vector Machine” se han configurado dos redes o sistemas con los parámetros mostrados en la tabla 3.1.

Posteriormente se han obtenido los resultados mostrados en la tabla 3.2 para cada sistema configurado respectivamente y junto a dichos resultados, se han obtenido las gráficas mostradas en la figura 3.3.

Red	Tree Depth	Lambda W	Lambda Theta	Lambda Theta Prime	Sigma	Number of Iterations
Azul:	3	0.1	0.01	0.01	1	15000
Roja:	6	0.057241	0.064953	0.05212	0.970426	10388

Tabla 3.4. Configuración de los modelos del bloque 4 para el melanoma

Red	Accuracy	Recall	Precision	F1 Score	AUC
Azul:	0.739	0.670	0.707	0.688	0.804
Roja:	0.721	0.639	0.689	0.663	0.780

Tabla 5.6. Resultados de los modelos del bloque 4 para el melanoma

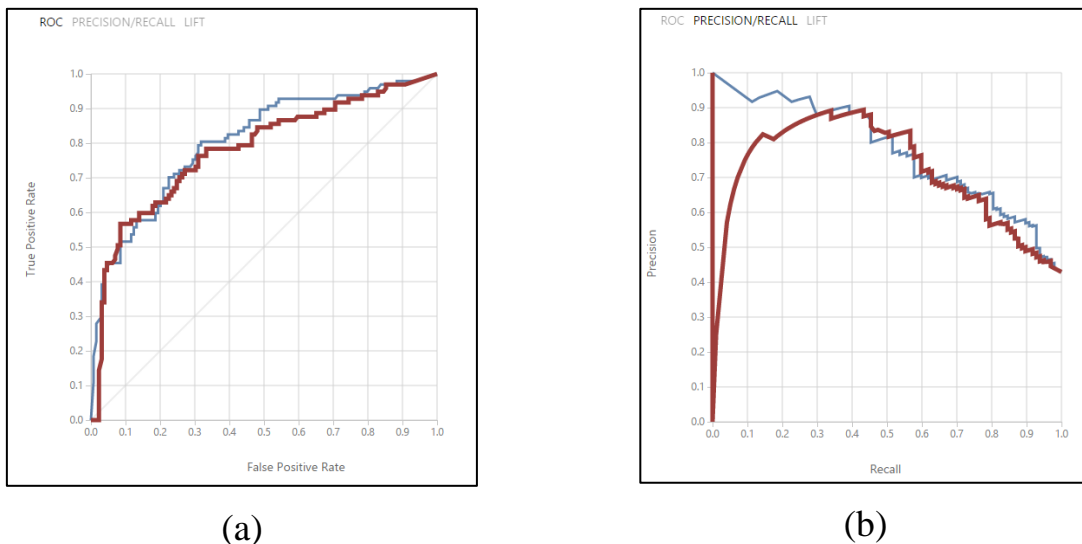


Figura 3.7. Gráficas de los resultados del bloque 4 para el melanoma (a) Curva ROC (b) Curva P/R

En esta comparación sale vencedora la red Azul por ofrecer mejores valores tanto de accuracy como de recall.

Para el análisis del bloque 2 correspondiente a los datos de melanoma con el modelo “Two-Class Support Vector Machine” se han configurado dos redes o sistemas con los parámetros mostrados en la tabla 3.3.

Posteriormente se han obtenido los resultados mostrados en la tabla 3.4 para cada sistema configurado respectivamente y junto a dichos resultados, se han obtenido las gráficas mostradas en la figura 3.4.

Red	Number of iterations	Lambda
Azul:	1	0.001
Roja:	17	0.000895

Tabla 8.9. Configuración de los modelos del bloque 4 para el melanoma

Red	Accuracy	Recall	Precision	F1 Score	AUC
Azul:	0.664	0.557	0.621	0.587	0.751
Roja:	0.690	0.629	0.642	0.635	0.751

Tabla 10.11. Resultados de los modelos del bloque 4 para el melanoma

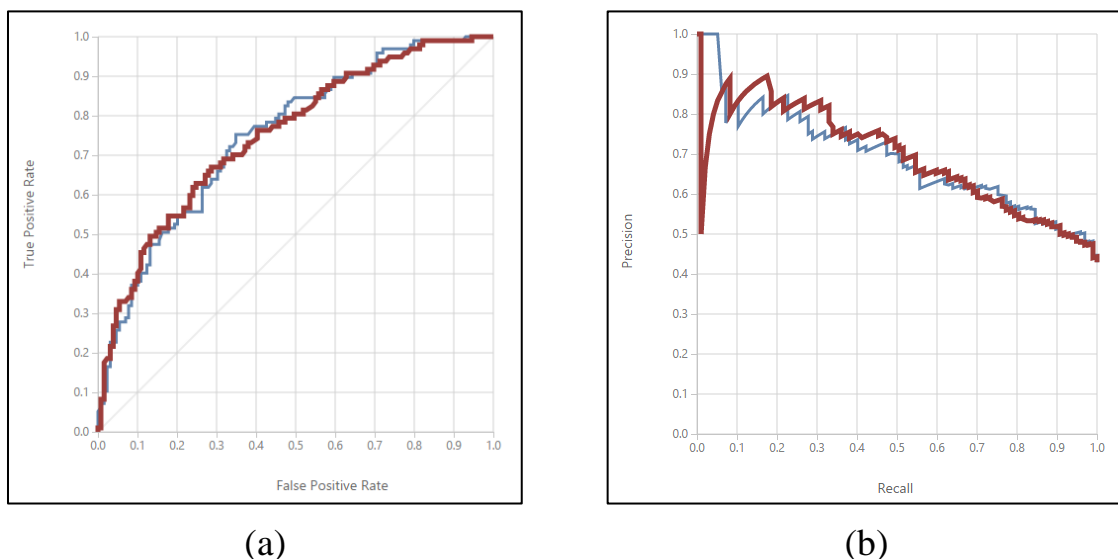


Figura 4.12. Gráficas de los resultados del bloque 4 para el melanoma (a) Curva ROC
(b) Curva P/R

En esta comparación sale vencedora la red Roja por ofrecer mejores valores tanto de accuracy como de recall

Para el análisis del bloque 3 correspondiente a los datos de melanoma con el modelo “Two-Class Neural Network” se han configurado dos redes o sistemas con los parámetros mostrados en la tabla 3.5.

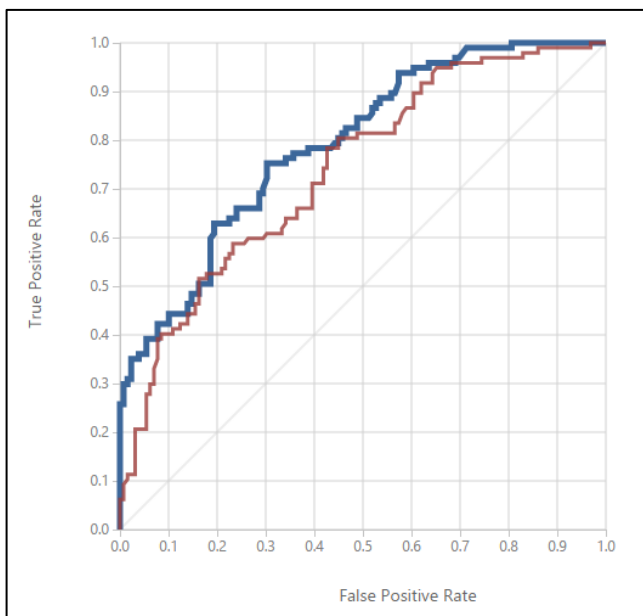
Posteriormente se han obtenido los resultados mostrados en la tabla 3.6 para cada sistema configurado respectivamente y junto a dichos resultados, se han obtenido las gráficas mostradas en la figura 3.5.

Red	Learning rate	Number of iterations
Azul:	0.1	100
Roja:	0.016576	54

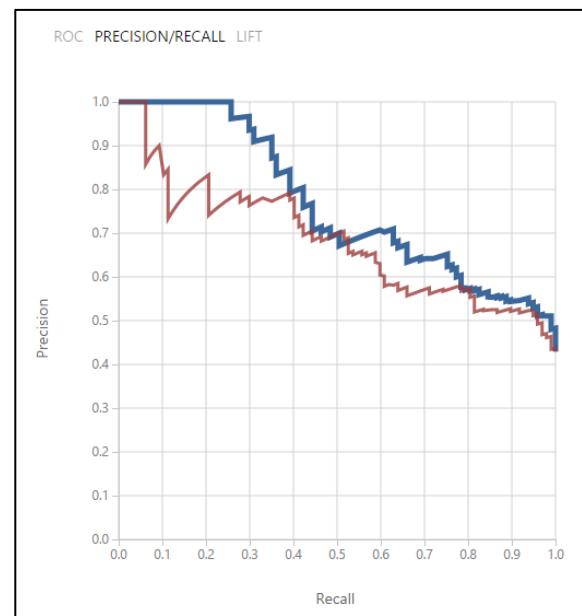
Tabla 13.14. Configuración de los modelos del bloque 4 para el melanoma

Red	Accuracy	Recall	Precision	F1 Score	AUC
Azul:	0.726	0.619	0.706	0.659	0.791
Roja:	0.681	0.588	0.640	0.613	0.739

Tabla 15.16. Resultados de los modelos del bloque 3 para el melanoma



(a)



(b)

Figura 5.17. Gráficas de los resultados del bloque 3 para el melanoma (a) Curva ROC (b) Curva P/R

En esta comparación sale vencedora la red Azul por ofrecer mejores valores tanto de accuracy como de recall.

Para el análisis del bloque 4 correspondiente a los datos de melanoma con el modelo “Two-Class Decision Jungle” se han configurado dos redes o sistemas con los parámetros mostrados en la tabla 3.7.

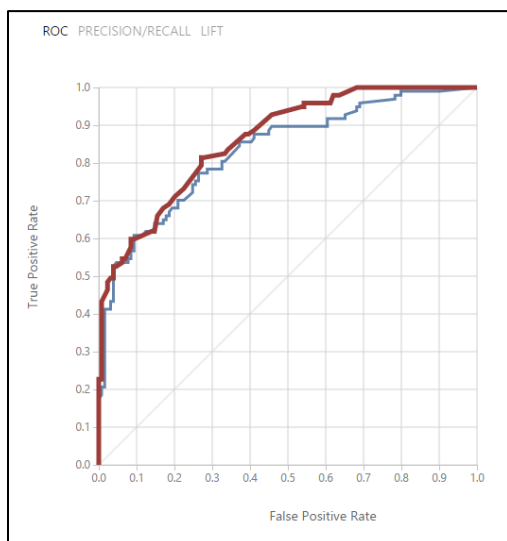
Posteriormente se han obtenido los resultados mostrados en la tabla 3.8 para cada sistema configurado respectivamente y junto a dichos resultados, se han obtenido las gráficas mostradas en la figura 3.6.

Red	Optimization steps	Maximum width of the decision DAGs	Maximum depth of the decision DAGs	Number of decision DAGs
Azul:	2048	128	32	8
Roja:	2235	150	104	30

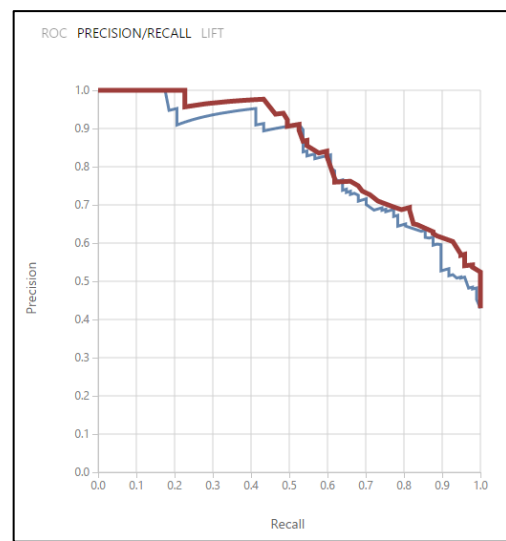
Tabla 18.19. Configuración de los modelos del bloque 4 para el melanoma

Red	Accuracy	Recall	Precision	F1 Score	AUC
Azul:	0.739	0.722	0.686	0.704	0.832
Roja:	0.761	0.691	0.736	0.713	0.862

Tabla 20.21. Resultados de los modelos del bloque 4 para el melanoma



(a)



(b)

Figura 6.22. Gráficas de los resultados del bloque 4 para el melanoma (a) Curva ROC (b) Curva P/R

En esta comparación la red Azul tiene un Recall mayor a la red Roja en 0,032 puntos, no obstante en el parámetro de Accuracy gana ventaja la red Roja por 0,022 puntos. Adicionalmente el parámetro AUC de la curva roja es mejor por lo tanto y teniendo en cuenta que la diferencia de Recall no es significativa escogemos a la red roja como vencedora

Una vez realizada la búsqueda de las 4 mejores configuraciones para cada modelo seleccionado para este análisis se procede a comparar dichas redes unas con otras en la tabla 3.9

Red	Accuracy	Recall	Precision	F1 Score	AUC
B1: Deep SVM.	0.739	0.670	0.707	0.688	0.804
B2: SVM	0.690	0.629	0.642	0.629	0.751
B3: Neural Network	0.726	0.619	0.706	0.659	0.791
B4: Decision Jungle	0.761	0.691	0.736	0.713	0.862

Tabla 23.24. Comparación final de modelos sobre el conjunto de entrenamiento y validación del melanoma

Estos resultados se traducen en que el sistema que ha obtenido mejores resultados de validación es el procedente del bloque 4, un modelo “Two-Class Decision Jungle”.

3.2.2 Cáncer de pulmón

Para el análisis del bloque 1 correspondiente a los datos de melanoma con el modelo “Two-Class Locally-Deep Support Vector Machine” se han configurado dos redes o sistemas con los parámetros mostrados en la tabla 3.10.

Posteriormente se han obtenido los resultados mostrados en la tabla 3.11 para cada sistema configurado respectivamente y junto a dichos resultados, se han obtenido las gráficas mostradas en la figura 3.7.

Red	Tree Depth	Lambda W	Lambda Theta	Lambda Theta Prime	Sigma	Number of Iterations
Azul:	3	0.1	0.01	0.01	1	15000
Roja:	6	0.057241	0.064953	0.05212	0.970426	10388

Tabla 25.26. Configuración de los modelos del bloque 1 para el cáncer de pulmón

Red	Accuracy	Recall	Precision	F1 Score	AUC
Azul:	0.883	0.813	0.891	0.850	0.916
Roja:	0.851	0.801	0.828	0.814	0.946

Tabla 27.28. Resultados de los modelos del bloque 1 para el cáncer de pulmón

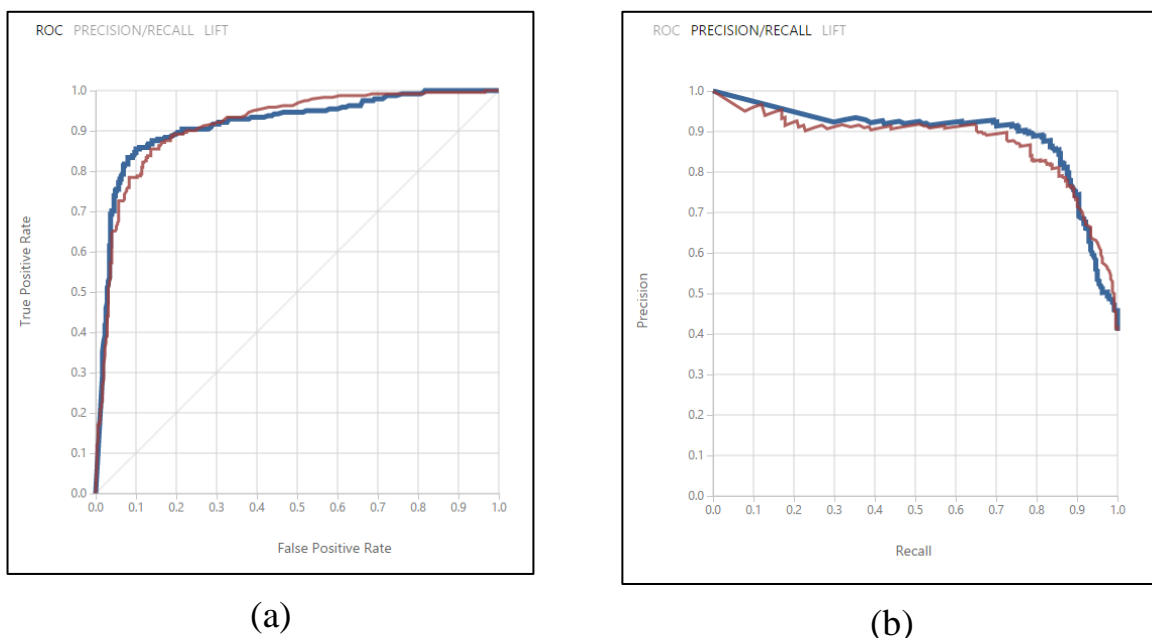


Figura 7.29. Gráficas de los resultados del bloque 1 para el cáncer de pulmón (a) Curva ROC (b) Curva P/R

En esta comparación sale vencedora la red Azul por ofrecer mejores valores tanto de accuracy como de recall.

Para el análisis del bloque 2 correspondiente a los datos de melanoma con el modelo “Two-Class Support Vector Machine” se han configurado dos redes o sistemas con los parámetros mostrados en la tabla 3.12.

Posteriormente se han obtenido los resultados mostrados en la tabla 3.13 para cada sistema configurado respectivamente y junto a dichos resultados, se han obtenido las gráficas mostradas en la figura 3.8.

Red	Number of iterations	Lambda
Azul:	1	0.001
Roja:	17	0.000895

Tabla 30.31. Configuración de los modelos del bloque 2 para el cáncer de pulmón

Red	Accuracy	Recall	Precision	F1 Score	AUC
Azul:	0.770	0.689	0.731	0.7709	0.824
Roja:	0.787	0.710	0.753	0.731	0.836

Tabla 32.33. Resultados de los modelos del bloque 2 para el cáncer de pulmón

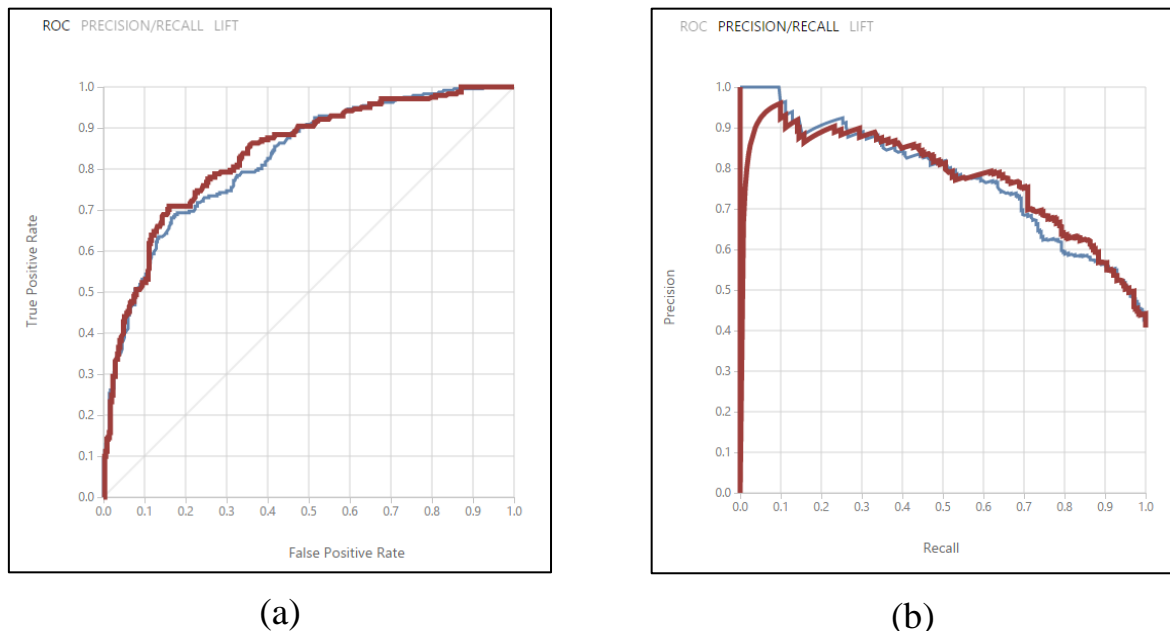


Figura 8.34. Gráficas de los resultados del bloque 2 para el cáncer de pulmón (a) Curva ROC (b) Curva P/R

En esta comparación sale vencedora la red Roja por ofrecer mejores valores tanto de accuracy como de recall.

Para el análisis del bloque 3 correspondiente a los datos de melanoma con el modelo “Two-Class Neural Network” se han configurado dos redes o sistemas con los parámetros mostrados en la tabla 3.14.

Posteriormente se han obtenido los resultados mostrados en la tabla 3.15 para cada sistema configurado respectivamente y junto a dichos resultados, se han obtenido las gráficas mostradas en la figura 3.9.

Red	Learning rate	Number of iterations
Azul:	0.1	100
Roja:	0.037861	129

Tabla 35.36. Configuración de los modelos del bloque 3 para el cáncer de pulmón

Red	Accuracy	Recall	Precision	F1 Score	AUC
Azul:	0.883	0.842	0.868	0.855	0.934
Roja:	0.873	0.817	0.864	0.840	0.931

Tabla 37.38. Resultados de los modelos del bloque 3 para el cáncer de pulmón

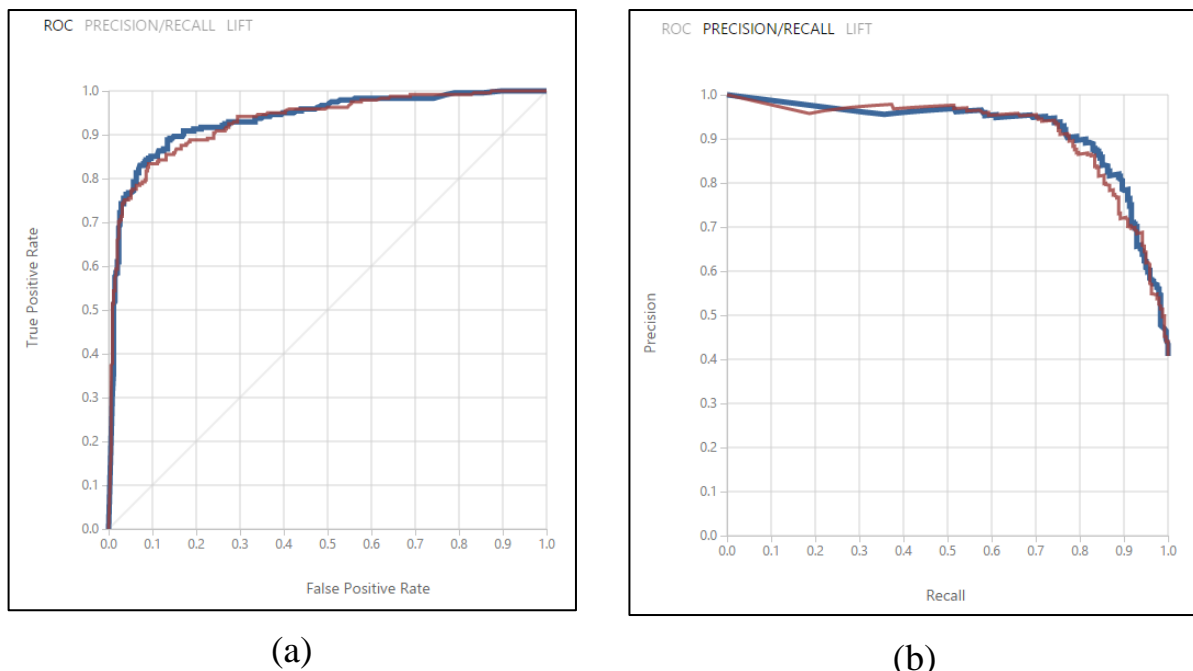


Figura 9.39. Gráficas de los resultados del bloque 3 para el cáncer de pulmón (a) Curva ROC (b) Curva P/R

En esta comparación sale vencedora la red Azul por ofrecer mejores valores tanto de accuracy como de recall.

Para el análisis del bloque 4 correspondiente a los datos de melanoma con el modelo “Two-Class Decision Jungle” se han configurado dos redes o sistemas con los parámetros mostrados en la tabla 3.16.

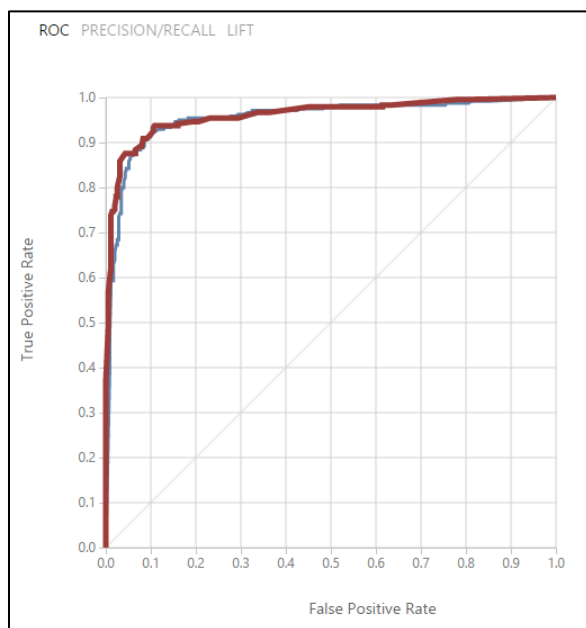
Posteriormente se han obtenido los resultados mostrados en la tabla 3.17 para cada sistema configurado respectivamente y junto a dichos resultados, se han obtenido las gráficas mostradas en la figura 3.10.

Red	Optimization steps	Maximum width of the decision DAGs	Maximum depth of the decision DAGs	Number of decision DAGs
Azul:	2048	128	32	8
Roja:	2235	150	104	30

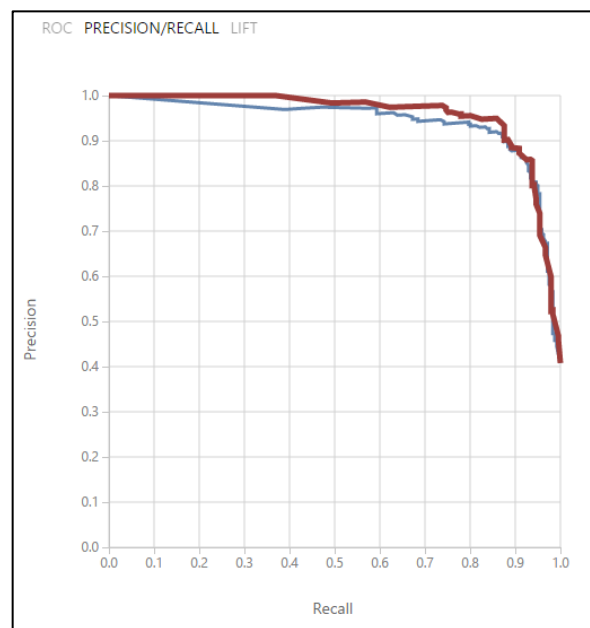
Tabla 40.41. Configuración de los modelos del bloque 4 para el cáncer de pulmón

Red	Accuracy	Recall	Precision	F1 Score	AUC
Azul:	0.907	0.834	0.931	0.880	0.952
Roja:	0.924	0.876	0.934	0.904	0.961

Tabla 42.43. Resultados de los modelos del bloque 4 para el cáncer de pulmón



(a)



(b)

Figura 10.44. Gráficas de los resultados del bloque 4 para el cáncer de pulmón (a) Curva ROC (b) Curva P/R

En esta comparación sale vencedora la red Roja por ofrecer mejores valores tanto de accuracy como de recall.

Una vez realizada la búsqueda de las 4 mejores configuraciones para cada modelo seleccionado para este análisis se procede a comparar dichas redes unas con otras en la tabla 3.18

Red	Accuracy	Recall	Precision	F1 Score	AUC
B1: Deep SVM.	0.883	0.813	0.891	0.850	0.916
B2: SVM	0.787	0.710	0.753	0.753	0.836
B3: Neural Network	0.883	0.842	0.868	0.855	0.934
B4: Decision Jungle	0.924	0.876	0.934	0.904	0.961

Tabla 45.46. Comparación final de modelos sobre el conjunto de entrenamiento y validación del cáncer de pulmón

Estos resultados se traducen en que el sistema que ha obtenido mejores resultados de validación es el procedente del bloque 4, un modelo “Two-Class Decision Jungle”.

3.2.3 Cáncer de mama

Para el análisis del bloque 1 correspondiente a los datos de melanoma con el modelo “Two-Class Locally-Deep Support Vector Machine” se han configurado dos redes o sistemas con los parámetros mostrados en la tabla 3.19.

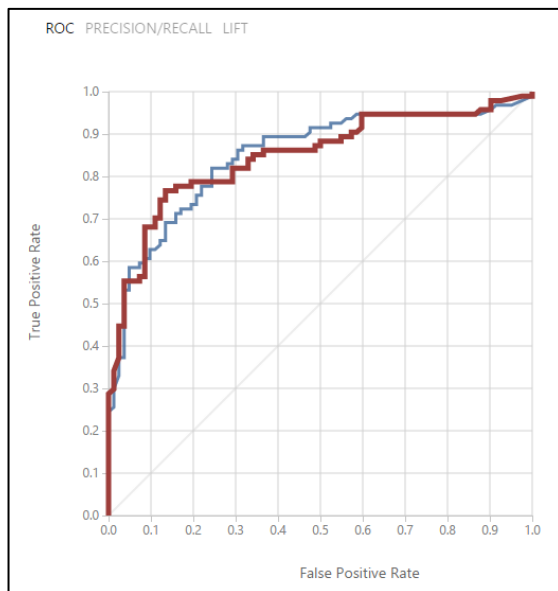
Posteriormente se han obtenido los resultados mostrados en la tabla 3.20 para cada sistema configurado respectivamente y junto a dichos resultados, se han obtenido las gráficas mostradas en la figura 3.11.

Red	Tree Depth	Lambda W	Lambda Theta	Lambda Theta Prime	Sigma	Number of Iterations
Azul:	3	0.1	0.01	0.01	1	15000
Roja:	6	0.057241	0.024971	0.05212	0.970426	10388

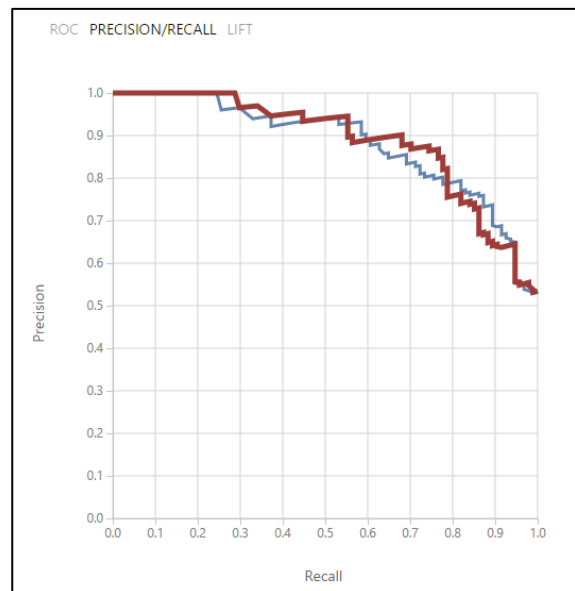
Tabla 47.48. Configuración de los modelos del bloque 1 para el cáncer de mama

Red	Accuracy	Recall	Precision	F1 Score	AUC
Azul:	0.767	0.777	0.785	0.781	0.849
Roja:	0.773	0.787	0.787	0.787	0.847

Tabla 49.50. Resultados de los modelos del bloque 1 para el cáncer de mama



(a)



(b)

Figura 11.51. Gráficas de los resultados del bloque 1 para el cáncer de mama (a) Curva ROC (b) Curva P/R

En esta comparación sale vencedora la red Roja por ofrecer mejores valores tanto de accuracy como de recall.

Para el análisis del bloque 2 correspondiente a los datos de melanoma con el modelo “Two-Class Support Vector Machine” se han configurado dos redes o sistemas con los parámetros mostrados en la tabla 3.21.

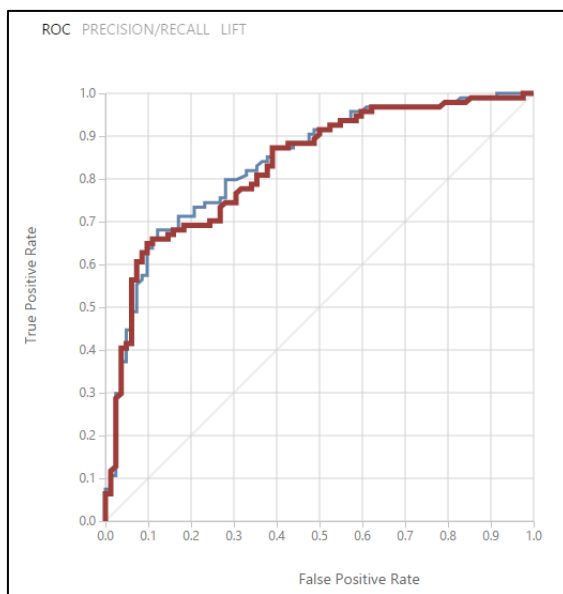
Posteriormente se han obtenido los resultados mostrados en la tabla 3.22 para cada sistema configurado respectivamente y junto a dichos resultados, se han obtenido las gráficas mostradas en la figura 3.12.

Red	Number of iterations	Lambda
Azul:	1	0.001
Roja:	6	0.007497

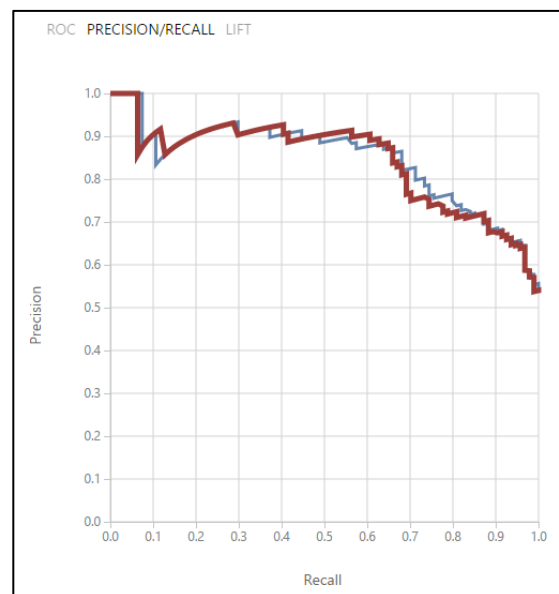
Tabla 52.53. Configuración de los modelos del bloque 2 para el cáncer de mama

Red	Accuracy	Recall	Precision	F1 Score	AUC
Azul:	0.756	0.787	0.763	0.775	0.838
Roja:	0.733	0.809	0.724	0.764	0.830

Tabla 54.55. Resultados de los modelos del bloque 2 para el cáncer de mama



(a)



(b)

Figura 12.56. Gráficas de los resultados del bloque 2 para el cáncer de mama (a) Curva ROC (b) Curva P/R

En esta comparación la red Roja tiene un Recall mayor a la red Azul en 0,022 puntos, no obstante en el parámetro de Accuracy gana ventaja la red Azul por 0,023 puntos. Adicionalmente el parámetro AUC de la curva azul es mejor por lo tanto y teniendo en cuenta que la diferencia de Recall no es significativa escogemos a la red azul como vencedora

Para el análisis del bloque 3 correspondiente a los datos de melanoma con el modelo “Two-Class Neural Network” se han configurado dos redes o sistemas con los parámetros mostrados en la tabla 3.23.

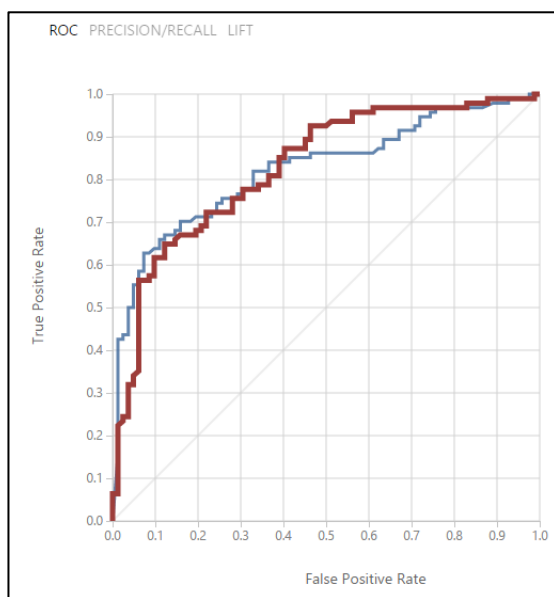
Posteriormente se han obtenido los resultados mostrados en la tabla 3.24 para cada sistema configurado respectivamente y junto a dichos resultados, se han obtenido las gráficas mostradas en la figura 3.13.

Red	Learning rate	Number of iterations
Azul:	0.1	100
Roja:	0.016576	54

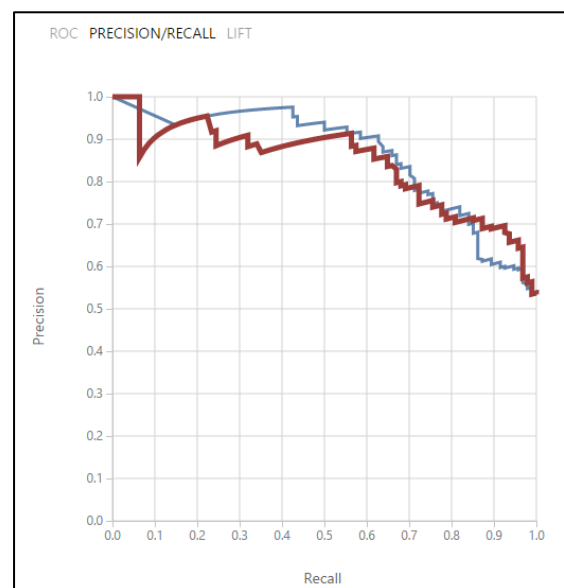
Tabla 57.58. Configuración de los modelos del bloque 3 para el cáncer de mama

Red	Accuracy	Recall	Precision	F1 Score	AUC
Azul:	0.744	0.809	0.738	0.772	0.822
Roja:	0.727	0.830	0.709	0.765	0.828

Tabla 59.604. Resultados de los modelos del bloque 3 para el cáncer de mama



(a)



(b)

Figura 13.61. Gráficas de los resultados del bloque 3 para el cáncer de mama (a) Curva ROC (b) Curva P/R

En esta comparación la red Roja tiene un Recall mayor a la red Azul en 0,021 puntos, no obstante en el parámetro de Accuracy gana ventaja la red Azul por 0,017 puntos. Adicionalmente el parámetro AUC de la curva roja es mejor por lo tanto y teniendo en cuenta que la diferencia de Accuracy no es significativa escogemos a la red roja como vencedora.

Para el análisis del bloque 4 correspondiente a los datos de melanoma con el modelo “Two-Class Decision Jungle” se han configurado dos redes o sistemas con los parámetros mostrados en la tabla 3.25.

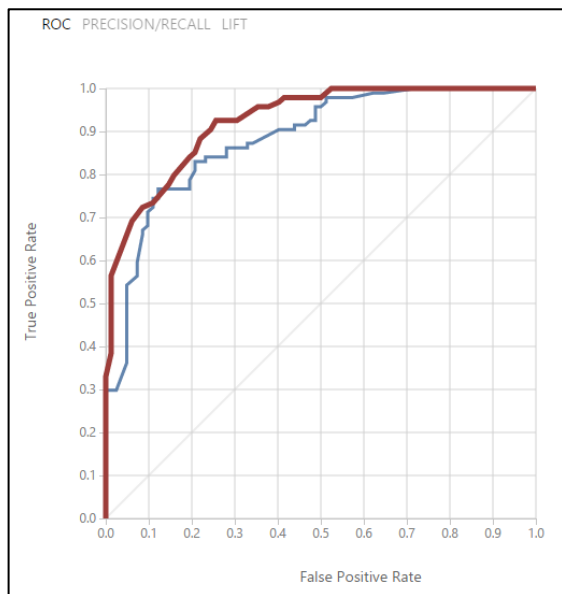
Posteriormente se han obtenido los resultados mostrados en la tabla 3.26 para cada sistema configurado respectivamente y junto a dichos resultados, se han obtenido las gráficas mostradas en la figura 3.14.

Red	Optimization steps	Maximum width of the decision DAGs	Maximum depth of the decision DAGs	Number of decision DAGs
Azul:	2048	128	32	8
Roja:	16384	8	128	32

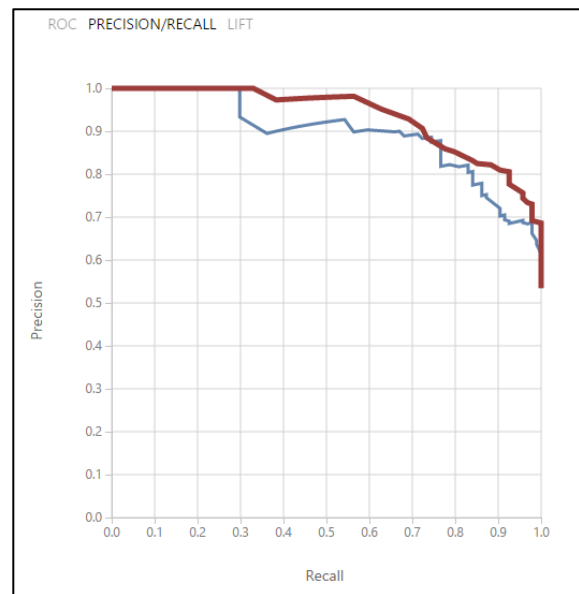
Tabla 62.63. Configuración de los modelos del bloque 4 para el cáncer de mama

Red	Accuracy	Recall	Precision	F1 Score	AUC
Azul:	0.790	0.840	0.782	0.810	0.882
Roja:	0.835	0.904	0.810	0.854	0.924

Tabla 64.656. Resultados de los modelos del bloque 4 para el cáncer de mama



(a)



(b)

Figura 14.66. Gráficas de los resultados del bloque 4 para el cáncer de mama (a) Curva ROC (b) Curva P/R

En esta comparación sale vencedora la red Roja por ofrecer mejores valores tanto de accuracy como de recall.

Una vez realizada la búsqueda de las 4 mejores configuraciones para cada modelo seleccionado para este análisis se procede a comparar dichas redes unas con otras en la tabla 3.27.

Red	Accuracy	Recall	Precision	F1 Score	AUC
B1: Deep SVM.	0.773	0.787	0.787	0.787	0.847
B2: SVM	0.756	0.787	0.763	0.775	0.838
B3: Neural Network	0.727	0.830	0.709	0.765	0.828
B4: Decision Jungle	0.790	0.840	0.782	0.810	0.882

Tabla 67.68. Comparación final de modelos sobre el conjunto de entrenamiento y validación del cáncer de mama

Estos resultados se traducen en que el sistema que ha obtenido mejores resultados de validación es el procedente del bloque 4, un modelo “Two-Class Decision Jungle”.

3.3 Estudio de los sistemas bajo el subconjunto de test

Los modelos obtenidos en los apartados anteriores serán puestos a prueba en esta sección bajo condiciones reales usando para ello el subconjunto de datos de test. Este subconjunto no ha sido usado antes para entrenar el modelo por lo que son datos que no conoce y por lo tanto servirán para obtener unos porcentajes de aciertos más próximos a la realidad.

Para este estudio se utiliza el esquema de módulos de Azure mostrado en la figura 3.15.

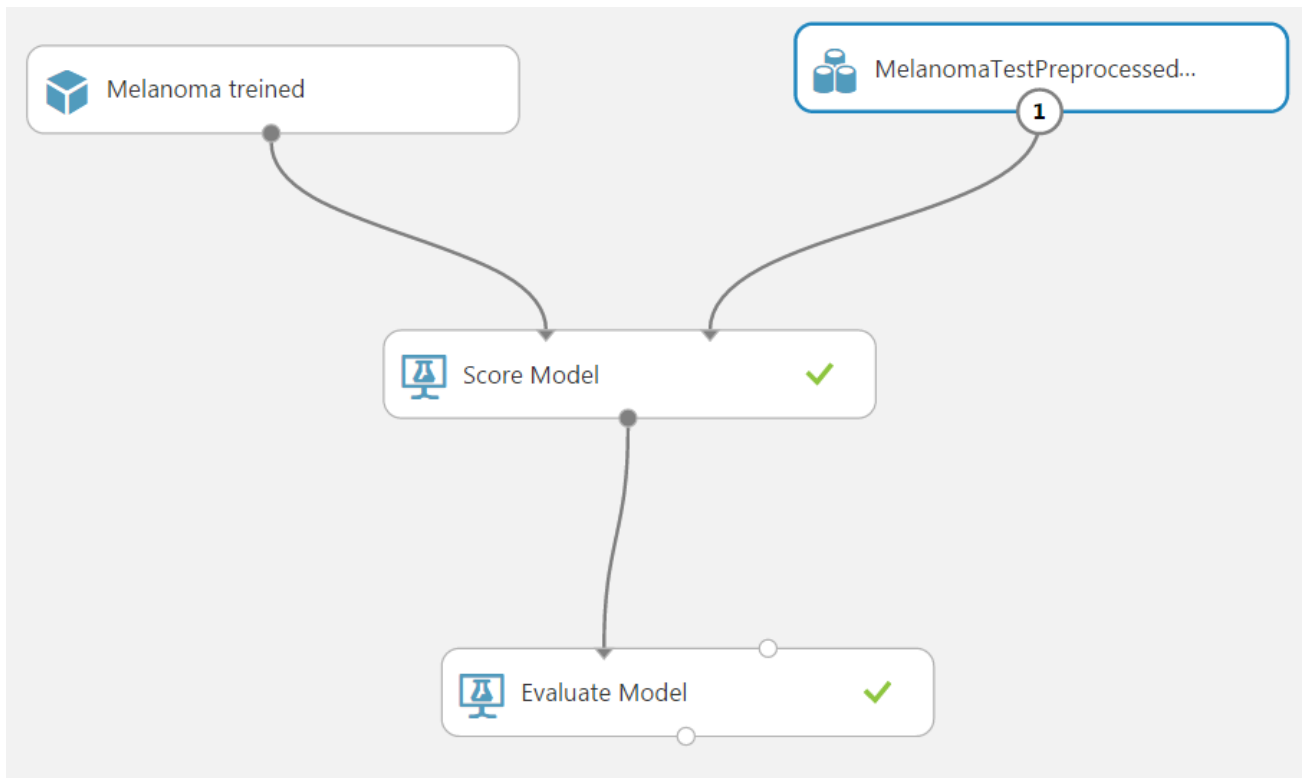


Figura 15.69. Esquema de módulos subconjunto de test

3.3.1 Melanoma

El mejor modelo hallado sobre el conjunto de entrenamiento y validación para los datos de melanoma ha sido el Decision Jungle. Es con este modelo con el que se ha hecho el estudio bajo el subconjunto de datos de test y cuyos resultados se muestran en la figura 3.16 y la tabla 3.28.

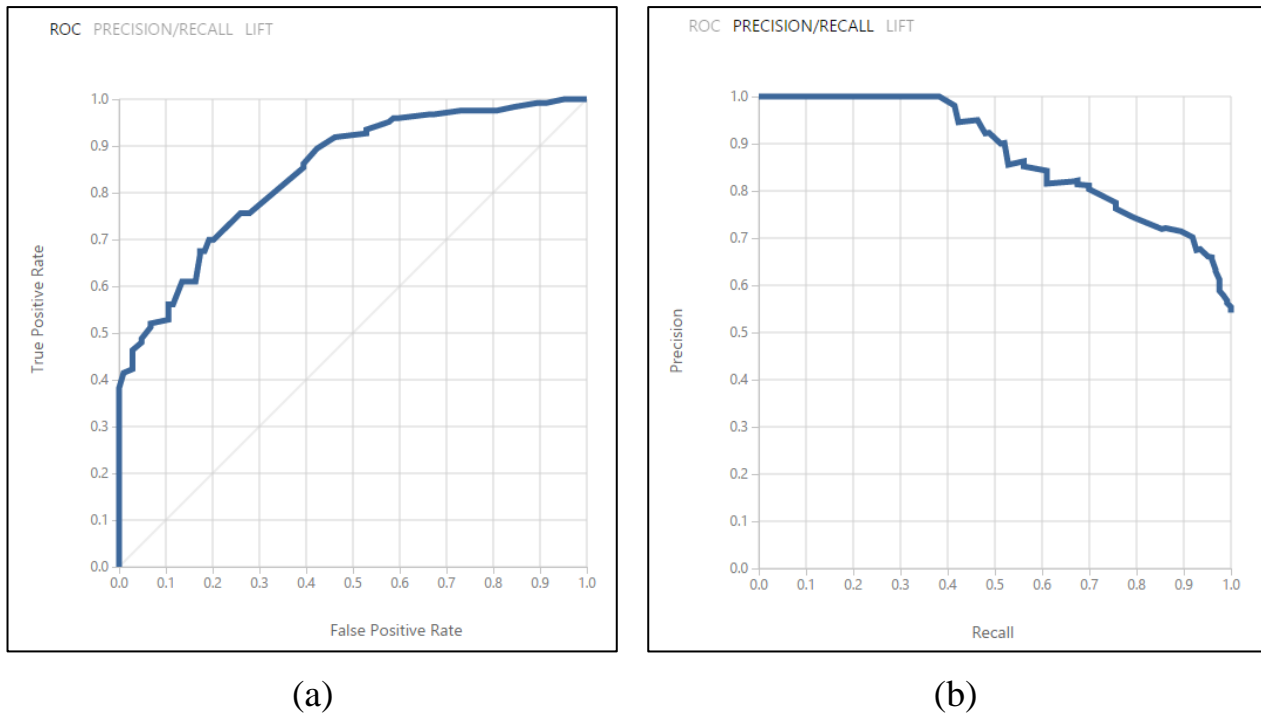


Figura 16.70. Gráficas del mejor modelo para el conjunto de test de los datos de melanoma (a) Curva ROC (b) Curva P/R

Modelo	Accuracy	Recall	Precision	F1 Score	AUC
Azure decision Jungle	0.749	0.699	0.811	0.751	0.842

Tabla 71.72. Resultados del mejor modelo para el conjunto de test de los datos de melanoma

3.3.2 Cáncer de pulmón

El mejor modelo hallado sobre el conjunto de entrenamiento y validación para los datos de cáncer de pulmón ha sido el Decision Jungle. Es con este modelo con el que se ha hecho el estudio bajo el subconjunto de datos de test y cuyos resultados se muestran en la figura 3.17 y la tabla 3.29.

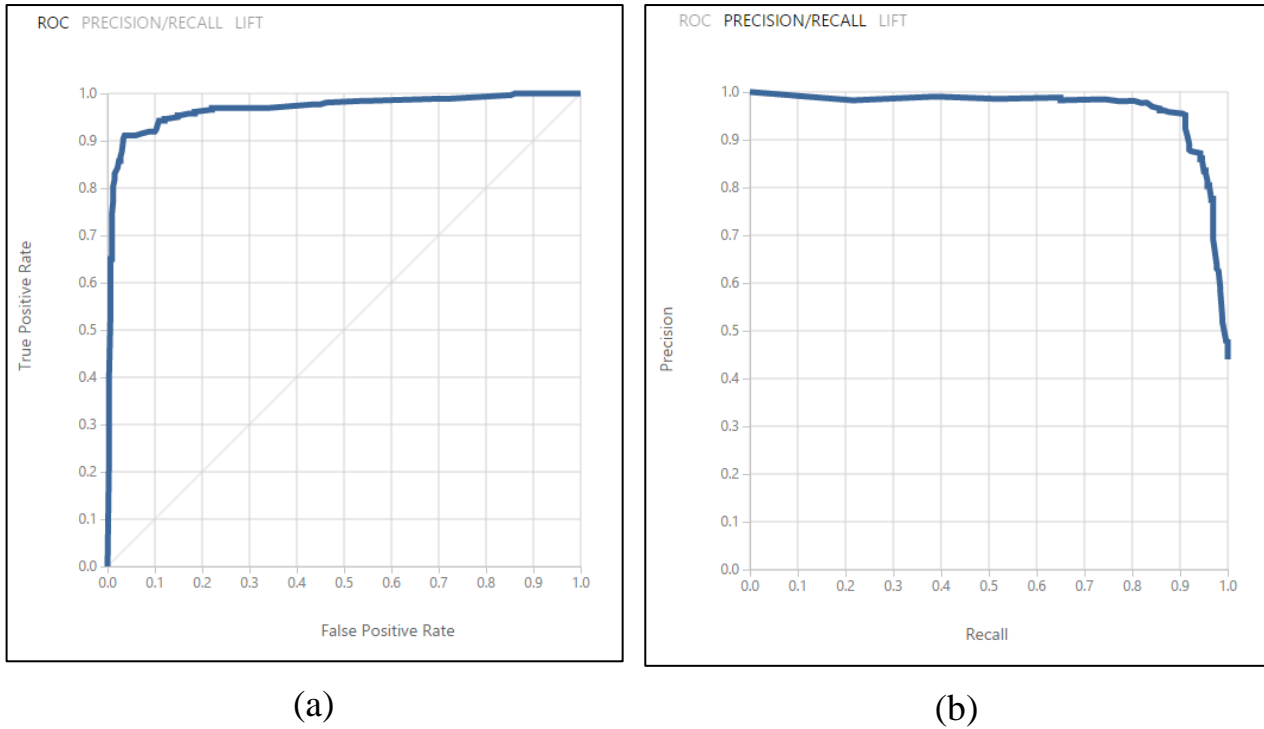


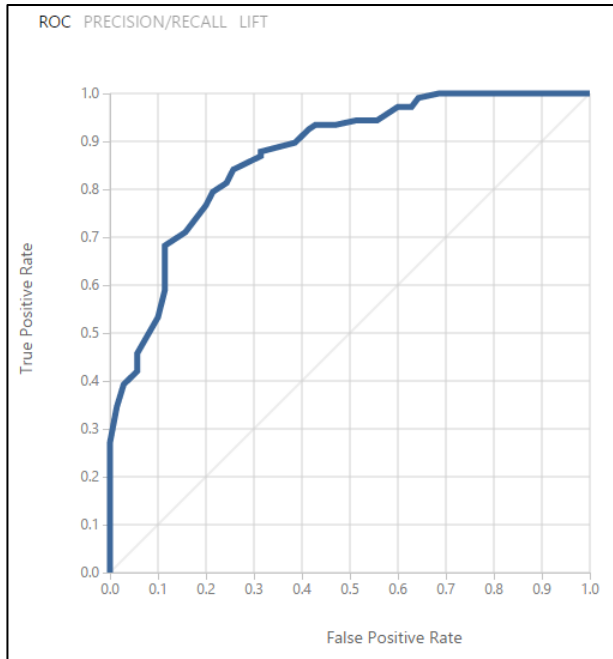
Figura 17.73. Gráficas del mejor modelo para el conjunto de test de los datos de cáncer de pulmón (a) Curva ROC (b) Curva P/R

Modelo	Accuracy	Recall	Precision	F1 Score	AUC
Azure decision Jungle	0.941	0.912	0.952	0.931	0.967

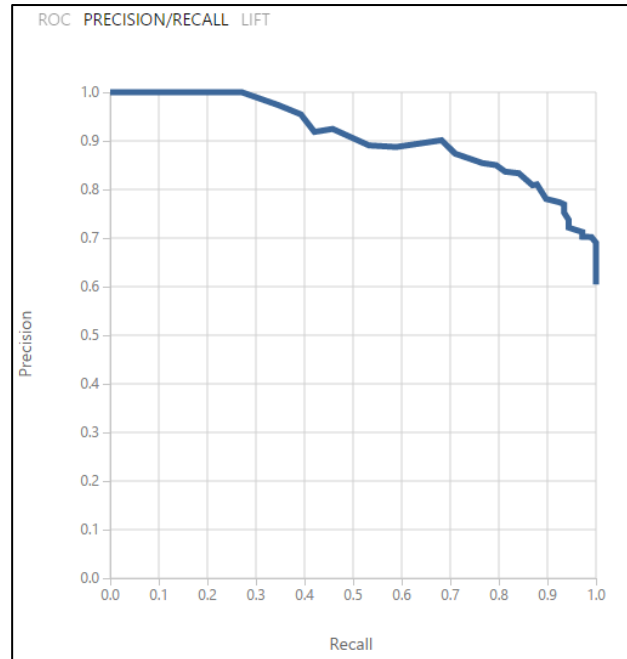
Tabla 74.75. Resultados del mejor modelo para el conjunto de test de los datos de cáncer de pulmón

3.3.3 Cáncer de mama

El mejor modelo hallado sobre el conjunto de entrenamiento y validación para los datos de cáncer de mama ha sido el Decision Jungle. Es con este modelo con el que se ha hecho el estudio bajo el subconjunto de datos de test y cuyos resultados se muestran en la figura 3.18 y la tabla 3.30.



(a)



(b)

Figura 18.76. Gráficas del mejor modelo para el conjunto de test de los datos de cáncer de pulmón (a) Curva ROC (b) Curva P/R

Modelo	Accuracy	Recall	Precision	F1 Score	AUC
Azure decision Jungle	0.797	0.869	0.809	0.838	0.870

Tabla 77.78. Resultados del mejor modelo para el conjunto de test de los datos de cáncer de mama

3.4 Comparación de resultados con HEAL

Hecho el estudio y obtenidos los resultados finales se procede a la comparación con los resultados obtenidos por el HEAL sobre el mismo conjunto de datos.

Importante recordar que el parámetro “sensitive” que estudia HEAL es el mismo que hemos denominado hasta ahora como “recall” a lo largo del estudio. Se usará este parámetro junto con el “accuracy” y la especificidad para medir y comparar ambos resultados.

También añadir que para enriquecer el análisis de los resultados se ha añadido en las respectivas tablas comparativas un apartado con los resultados obtenidos en el subconjunto de test por el alumno Alberto Fariña Barrera en el desarrollo de su trabajo de fin de grado, análogo a este. En dicho trabajo Alberto ha usado el mismo conjunto de datos que en este estudio a excepción de que los suyos si se encontraban normalizados. PyBrain ha sido la herramienta que ha usado para desarrollar sus pruebas y Counter-Propagation y Back-Propagation los modelos con los que ha obtenido mejores resultados.

3.4.1 Melanoma

En las tablas 3.31 y 3.32 se muestra la relación de los resultados obtenidos en este trabajo de fin de grado con los obtenidos por HEAL con los datos de Melanoma.

Resultados	Entrenamiento y validación			
	Accuracy	Recall/Sensitive	Especificidad	Filas eliminadas (%)
Azure Decision Jungle	0.761	0.691	0.8139	0.1171
HEAL	0.8320	0.8024	0.8577	0

Tabla 79.80. Comparación melanoma con resultados de HEAL, subconjunto de entrenamiento y validación

Resultados	Testeo			
	Accuracy	Recall/Sensitive	Especificidad	Filas eliminadas
Azure Decision Jungle	0.749	0.699	0.8076	0.1171
HEAL	0.7315	0.6667	0.7876	0
Back-Propagation	0.7310	0.7640	0.6920	0.1171

Tabla 81.82. Comparación melanoma con resultados de HEAL, subconjunto de tests

En el caso del análisis realizado para la detección de melanoma se concluye que el estudio realizado por el HEAL ha obtenido mejores resultados con el subconjunto de datos de entrenamiento que el estudio realizado para este trabajo de fin de grado. No obstante sobre el subconjunto de datos de testeo se han obtenido mejores resultados con el estudio realizado en este trabajo. No obstante, considerando el porcentaje de

filas eliminadas por el preprocesado, que en nuestro sistema podríamos considerar esos casos como desconocidos, los resultados de HEAL son mucho mejores

3.4.2 Cáncer de pulmón

En las tablas 3.33 y 3.34 se muestra la relación de los resultados obtenidos en este trabajo de fin de grado con los obtenidos por HEAL con los datos de cáncer de pulmón.

	Entrenamiento y validación			
Resultados	Accuracy	Recall/Sensitive	Especificidad	Filas eliminadas (%)
Azure Decision Jungle	0.924	0.876	0.9571	0.1557
HEAL	0.9293	0.8896	0.9583	0

Tabla 83.84. Comparación cáncer de pulmón con resultados de HEAL, subconjunto de entrenamiento y validación

	Testeo			
Resultados	Accuracy	Recall/Sensitive	Especificidad	Filas eliminadas (%)
Azure Decision Jungle	0.941	0.912	0.9637	0.1557
HEAL	0.8730	0.7761	0.9437	0
Back-Propagation	0.7310	0.8770	0.6160	0.1557

Tabla 85.86. Comparación cáncer de pulmón con resultados de HEAL, subconjunto de tests

En cuanto al cáncer de pulmón los resultados globales se asemejan a los obtenidos para el melanoma. En el subconjunto de datos de entrenamiento el HEAL obtiene mejores resultados pero en el subconjunto de testeo es el estudio desarrollado para este trabajo de fin de grado el que obtiene un mejor rendimiento. No obstante, considerando el porcentaje de filas eliminadas por el preprocesado, que en nuestro sistema podríamos considerar esos casos como desconocidos, los resultados de HEAL son mucho mejores.

3.4.3 Cáncer de mama

En las tablas 3.35 y 3.36 se muestra la relación de los resultados obtenidos en este trabajo de fin de grado con los obtenidos por HEAL con los datos de cáncer de pulmón.

	Entrenamiento y validación			
Resultados	Accuracy	Recall/Sensitive	Especificidad	Filas eliminadas (%)
Azure Decision Jungle	0.790	0.840	0.7317	0.0127
HEAL	0.7705	0.9555	0.5525	0

Tabla 87.885. Comparación cáncer de mama con resultados de HEAL, subconjunto de entrenamiento y validación

	Testeo			
Resultados	Accuracy	Recall/Sensitive	Especificidad	Filas eliminadas (%)
Azure Decision Jungle	0.797	0.869	0.6857	0.0127
HEAL	0.7436	0.9188	0.5370	0
Counter-Propagation	0.644	0.822	0.371	0.0127

Tabla 89.906. Comparación cáncer de mama con resultados de HEAL, subconjunto de tests

Para el cáncer de mama se concluye que tanto el subconjunto de entrenamiento como el de testeo se han obtenido mejores índices de sensibilidad en los estudios realizados por el HEAL. Por otro lado se observa un accuracy mayor en el estudio realizado para este trabajo de fin de grado incluso incluyendo el porcentaje de filas eliminadas por el preprocesado.

Capítulo 4. Conclusiones y líneas futuras

Los objetivos marcados para este trabajo de fin de grado se han realizado con éxito, incluso los resultados finales se han acercado más de lo esperado a los obtenidos por HEAL.

Las líneas futuras para el desarrollo del proyecto son mejorar el preprocesado de datos en cuanto a normalizar los datos y comprobar si los resultados mejoran. También sería interesante hacer pruebas con el resto de algoritmos de clasificación que ofrece Azure ML para comparar los resultados obtenidos. Otra forma de mejorar el sistema creado es conseguir ampliar el número de datos de entrada para poder entrenar más el sistema.

En conclusión, la inteligencia artificial ha llegado al día a día de nuestras vidas para quedarse. Son muchos ya los ámbitos donde se encuentra y muchas más el número de puertas que nos abre de cara al futuro de la vida de la sociedad humana tal y como la entendemos.

Concretando más en el mundo de la medicina queda claro que un sistema similar al desarrollado para este estudio puede llegar a ser de utilidad para automatizar y apoyar a los profesionales del sector en tareas de diagnóstico y en un futuro quizás llegar a delegar esta labor por completo a un sistema inteligente.

Capítulo 5.

Summary and Conclusions

All targets setter for this match were successful achieved, even the results have been similar to the results of HEAL.

The future of the improving of the project could be in update the preprocess of the data base normalizing the data and checking if the result improve. Also is interesting to make tests with the rest of classifications algorithms of Azure ML. Another way to improve the system is getting more training data.

In conclusion, the artificial intelligence has arrived to our live for ever. There are many scopes where AI can be used and more the opportunities that it create in the lives of human society.

Approaching to the world of medicine it is clear a similar system like the one that has been developed for these research could become useful for automate and support doctors in diagnostic tasks. And in the future maybe, delegate this work totally for these systems.

Capítulo 6.

Presupuesto

6.1 Conceptos

Las horas dedicadas por un ingeniero informático se ha estimado en un total de 300, valoradas en 30 euros/hora resulta un total de 3000 euros. Esta es la división de horas:

- a) Seminarios: 15
- b) Estudio autónomo: 65
- c) Tareas y trabajos: 200
- d) Reuniones y tutorías: 20

Por otro lado se han estimado un total de 40 euros destinados a transporte público para realizar los seminarios y reuniones necesarias.

En la tabla 6.1 se muestran los gastos totales para cada concepto y para el total del proyecto.

Concepto	Coste
Horas Ingeniero Informático	9000 €
Transporte	40 €
Total	9040 €

Tabla 1. 2. Presupuesto.

Bibliografía

- [1] **Enlace GitHub al trabajo fin de grado.**
<https://github.com/etsiiull/azuremedicalnn>
- [2] **Inteligencia Artificial.** https://es.wikipedia.org/wiki/Inteligencia_artificial
- [3] **Redes Neuronales Artificiales, un enfoque práctico.** Pedro Isasi Viñuela e Inés M. Galván León – Ed. Pearson, 2004
- [4] **Redes Neuronales y Sistemas Borrosos.** Bonifacio Marín del Brío y Alfredo Sanz Molina – Ed. Ra-Ma, 2ª edición, 2001
- [5] **Google research.** <https://research.google.com/>
- [6] **IBM research.** <http://www.research.ibm.com/>
- [7] **Microsoft Azure Machine Learning.**
<https://azure.microsoft.com/es-es/services/machine-learning/>
- [8] **Twitter machine Learning.** <https://engineering.twitter.com/cortex>
- [9] **Inteligencia Artificial en medicina.**
<http://www.xataka.com/medicina-y-salud/cuando-el-doctor-robot-diagnostica-mejor-que-el-ser-humano>
- [10] **Babylon: IA para asesoramiento de pacientes.** <http://www.babylonhealth.com/>
- [11] **HEAL** <http://heal.heuristicslab.com/>
- [12] **Documentación NeuroLab.** <https://pythonhosted.org/neurolab/>
- [13] **Documentación NeuralNet.**
<https://cran.r-project.org/web/packages/neuralnet/neuralnet.pdf>
- [14] **Documentación PyBrain.**
<http://pybrain.org/docs/>
- [15] **Documentación Fast Artificial Neural Network.**
<http://libfann.github.io/fann/docs/files/fann-h.html>
- [16] **Documentación Tensor Flow.**
<https://www.tensorflow.org/>
- [17] **Documentación IBM SPSS Modeler.**
<http://www-03.ibm.com/software/products/es/spss-modeler>
- [18] **Documentación Weka.**
<http://www.cs.waikato.ac.nz/ml/weka/>

- [19] **Documentación Azure Machine Learning.**
<https://msdn.microsoft.com/en-us/library/azure/dn578280.aspx>
- [20] **Perceptrón.** <http://www.lcc.uma.es/~jmortiz/archivos/Tema4.pdf>
- [21] **SVM.**
<http://www.asepelt.org/ficheros/File/Anales/2003%20-%20Almeria/asepeltPDF/55.pdf>
- [22] **Neural Network.**
<https://msdn.microsoft.com/en-us/library/azure/dn905947.aspx>
- [23] **Deep SVM.**
<https://msdn.microsoft.com/en-us/library/azure/dn913070.aspx>
- [24] **Árbol de decisión.**
http://www.dmae.upct.es/~mcruiz/Telem06/Teoria/arbol_decision.pdf
- [25] **Computer Aided Systems Theory – Eurocast 2015.** Michael, Karin Zölzer, Stephan M. Winkler, Erwin Hopf, Herbert Stekel, Rupert Frechinger y Stefan Wagner, 2015. Capítulo: Increasing the Sensitivity of Cancer Predictors Using Confidence Based Ensemble Modeling. 350-358. 2015