



DETECCIÓN AUTOMATIZADA DEL GLIOBLASTOMA MULTIFORME

Memoria Final PBL



7 DE JUNIO DE 2021

AINHOA ARRUABARRENA ORTIZ
EDGAR AZPIAZU CRESPO

RESUMEN

El proyecto descrito en el documento se corresponde con el PBL (Project Based Learning) de título “Detección Automatizada del Glioblastoma Multiforme” desarrollado en la Universidad de Mondragón en el Máster de Tecnologías Biomédicas. Se presenta la implementación de un sistema automatizado, cuyos objetivos son la identificación y diagnóstico del Glioblastoma Multiforme en resonancias magnéticas mediante técnicas de procesamiento de imagen y el desarrollo de un modelo de inteligencia artificial capaz de realizar una predicción de las probabilidades de supervivencia del paciente. El documento está compuesto por un estado del arte, seguido por el diseño y la implementación de la solución. Se han analizado e identificado los biomarcadores de los pacientes y características de las regiones tumorales que más influyen en la enfermedad. Por último, se ha podido predecir la supervivencia a los 12 meses con un *Accuracy* del 0.67.

ABSTRACT

The project described in the document corresponds to the PBL (Project Based Learning) entitled "Detección Automatizada del Glioblastoma Multiforme" developed at the University of Mondragon in the Master of Biomedical Technologies. This report presents the implementation of an automated system whose objectives are the identification and diagnosis of Multiform Glioblastoma in magnetic resonances using image processing techniques and the development of an intelligence model capable of making a prediction of the patient's survival probabilities. The document is composed of a state of the art, followed by the design and the implementation of the solution. The biomarkers of the patients and characteristics of the tumour regions that most influence the disease have been analysed and identified. Finally, 12-month survival has been predicted with an *Accuracy* of 0.67.

LABURPENA

Dokumentu honetan azaltzen den proiektua Mondragon Unibertsitateko Teknologia Biomedikoen Masterrean garatu da PBL (Project Based Learning) metodologia jarraituz “Detección Automatizada del Glioblastoma Multiforme” izendapenarekin. Helburutzat irudi prozesaketa teknikak erabiliz erresonantzia magnetikoetan Glioblastoma Multiformeen identifikazioa duen proiektuaren inplementazioa aurkezten da dokumentu honetan. Gainera, beraien diagnostikoa egin nahi da alde batetik eta gaixoen biziraupen aukerak iragarriko dituen adimen artifizialeko eredu bat bestetik. Dokumentu honek literaturaren analisi bat aurkezten du, garatutako sistemaren diseinu eta inplementazioarekin batera. Gaixoen biomarkadoreak eta tumoreen irudien ezaugarri garrantzitsuenak aztertu eta identifikatu dira. Azkenik, gaixoei 12 hilabete bizirauteko dituzten aukerak 0.67-ko *Accuracy*-arekin iragarri dira.

ÍNDICE

1.	INTRODUCCIÓN	7
1.1.	Contexto.....	7
1.2.	Objetivos	9
1.3.	Planificación	9
1.3.1.	Diagrama de Gantt	11
2.	ESTADO DEL ARTE	12
2.1.	Glioblastoma multiforme.....	12
2.1.1.	Clasificación de los <i>gliomas</i>	12
2.1.2.	Adquisición de imágenes	12
2.1.3.	Proceso de segmentación	13
2.1.4.	Métodos de segmentación	15
2.2.	Técnicas de análisis	17
2.2.1.	Análisis de supervivencia	17
2.2.2.	Métodos de Machine Learning	19
2.3.	Detección de Glioblastoma mediante Machine Learning	29
2.3.1.	Técnicas tradicionales de Machine Learning	29
2.3.2.	Nuevas técnicas de Deep Learning	31
3.	DISEÑO DE LA SOLUCIÓN	33
3.1.	Requisitos técnicos	33
3.2.	Arquitectura	34
3.3.	Dataset	35
3.3.1.	Pre-procesamiento del dataset	35
3.4.	Herramientas	36
4.	DESARROLLO	37
4.1.	Pre-procesamiento de imágenes	37
4.1.1.	Estandarización de las escalas de intensidades	37
4.1.2.	Eliminación del cráneo	38
4.2.	Segmentación del <i>glioma</i> y del edema.....	41
4.2.1.	Mezcla de Gaussianas	41
4.2.2.	Reconstrucción morfológica de la escala de grises	42

4.2.3.	<i>Threshold</i> binario adaptativo	43
4.2.4.	Detección de la región con mayor intensidad media	43
4.2.5.	Operaciones morfológicas	43
4.2.6.	Validación de la segmentación	44
4.2.7.	Resultados de la segmentación	46
4.3.	Extracción de características	46
4.3.1.	Empleo de distintos tipos de escáneres	48
4.4.	Análisis de los factores de riesgo	48
4.4.1.	Factores de riesgo basados en metadatos	48
4.4.2.	Factores de riesgo basados en características de las resonancias magnéticas y metadatos	52
4.5.	Modelo de predicción de supervivencia.....	56
4.5.1.	Comparación de los modelos.....	57
5.	EVALUACIÓN	58
5.1.	Algoritmo para la eliminación del cráneo.....	58
5.1.1.	Dataset	58
5.1.2.	Resultados.....	58
5.2.	Modelo final de Machine Learning	59
6.	CONCLUSIONES	60
7.	LÍNEAS FUTURAS	62
8.	BIBLIOGRAFÍA	63
APÉNDICE A.	DATOS DE LA EVALUACIÓN.....	I
APÉNDICE B.	RESULTADOS DEL ANÁLISIS ESTÁTICO	II

ÍNDICE DE FIGURAS

Figura 1. Proceso del diagnóstico. Figura traducida de [8].	8
Figura 2: Diagrama de Gantt del proyecto.	11
Figura 3. Clasificación de los gliomas. Figura traducida de [11].	12
Figura 4. Tipos de RM para la detección de tumores. Figura traducida de [15].	13
Figura 5. Bloques principales utilizados en la segmentación. Figura traducida de [17].	14
Figura 6. Segmentación de tumor cerebral semiautomática [27].	16
Figura 7. Ejemplo de curva de supervivencia.	18
Figura 8. Ejemplo de aprendizaje no supervisado. Figura traducida de [41].	22
Figura 9: Arquitectura simple de una CNN, formada por 5 capas [49]	24
Figura 10. Matriz de confusión. Figura traducida de [16].	25
Figura 11. Curva PR para distintos valores de umbral [52].	26
Figura 12. Ejemplo de curva ROC [52].	27
Figura 13. Histogramas de la distribución de las características más predictivas de la esperanza de vida [65].	30
Figura 14. Flujo de trabajo del análisis radiómico [72].	31
Figura 15. CNN usada para la segmentación de tumores cerebrales [74].	32
Figura 16. Esquema de la arquitectura del diseño de la solución.	34
Figura 17. Parámetros correspondientes al histograma de una RM [83].	37
Figura 18. Comparación entre las RM originales y estandarizadas.	38
Figura 19. Algoritmo McStrip. Imagen traducida de [85].	39
Figura 20. Máscara del Nivel 1 obtenida en el algoritmo McStrip.	39
Figura 21. Máscara del Nivel 2 obtenida en el algoritmo McStrip.	40
Figura 22. Diagrama del algoritmo BSE.	40
Figura 23. Máscara del Nivel 3 obtenida en el algoritmo McStrip.	40
Figura 24. Máscara de Consenso obtenida en el algoritmo McStrip.	41
Figura 25: Algoritmo implementado para la segmentación del tumor.	41
Figura 26: Aproximación del histograma mediante la mezcla de Gaussianas [80].	42
Figura 27: Determinación de los h-domes de una imagen en escala de grises [82]	42
Figura 28: Proceso de segmentación.	43
Figura 29: Análisis de continuidad entre máscaras.	45
Figura 30: Resultados de la validación de las máscaras.	45
Figura 31: Resultado de la segmentación de un tumor.	46
Figura 32: Resultado de la segmentación de un edema.	46
Figura 33: Gráficas extraídas para la separación de clases del análisis de factores de riesgo con los metadatos.	50
Figura 34: Curvas de supervivencia obtenidas del análisis de factores de riesgo con los metadatos.	51
Figura 35: Gráficas extraídas para la separación de clases del análisis de factores de riesgo con los metadatos y características.	53
Figura 36: Curvas de supervivencia obtenidas del análisis de factores de riesgo con los metadatos y características extraídas.	55

Figura 37: Valores de las métricas Accuracy y LogLoss para los modelos creados.....	57
Figura 38: Matrices de confusión obtenidas en la validación de los modelos.....	59

ÍNDICE DE TABLAS

Tabla 1. Resultados de los algoritmos de aprendizaje supervisado [42].	21
Tabla 2. Información sobre el dataset	35
Tabla 3: Características extraídas para su posterior análisis.....	47
Tabla 4. Descripción de las variables disponibles para el análisis de factores de riesgo basado en los metadatos.....	48
Tabla 5. Resultados de relevancia por cada variable y modelo en el análisis de factores de riesgo con los metadatos.....	51
Tabla 6. Valores de la concordancia obtenidos para los modelos de supervivencia creados de los metadatos.....	52
Tabla 7: Resultados de relevancia por cada variable y modelo en el análisis de factores de riesgo con los metadatos y características.....	55
Tabla 8: Valores de la concordancia obtenidos para los modelos de supervivencia creados de los metadatos y características.....	56
Tabla 9. Media de las métricas obtenidas para la evaluación del algoritmo de extracción del cerebro.....	58

LISTA DE ACRÓNIMOS

ACC. *Accuracy*
ANN. *Artificial Neural Network*
AUC. *Area Under the Curve*
BD. *Base de Datos*
BSE. *Brain Surface Extractor*
CART. *Classification And Regression Trees*
CNN. *Redes neuronales convolucionales*
CR. *Radiografía Computarizada*
DL. *Deep Learning*
EDA. *Exploratory Data Analysis*
FLAIR. *Fluid Attenuated Inversion Recovery*
FN. *False Negative*
FNR. *False Negative Rate*
FP. *False Positive*
FPR. *False Positive Rate*
GBM. *Glioblastoma Multiforme*
GLCM. *Grey Level Co-occurrence Matrix*
IA. *Inteligencia Artificial*
IBSR. *Internet Brain Segmentation Repository*
ICA. *Independent Component Analysis*
IMC. *Índice de Masa Corporal*
LoG. *Laplacian of Gaussian*
ML. *Machine Learning*
NA. *Not Available*
NPV. *Negative Predictive Value*
OMS. *Organización Mundial de la Salud*
PCA. *Principal Component Analysis*
PMM. *Predictive Mean Matching*
PPV. *Positive Predictive Value*
RF. *Random Forest*
RM. *Resonancia Magnética*
ROC. *Receiver Operating Characteristic*
ROI. *Región de interés*
SNR. *Signal to Noise Ratio*
SOM. *Self-organizing Map*
SVM. *Support Vector Machines*
SVR. *Support Vector Regression*
SVRF. *Support Vector Random Field*
TAC. *Tomografía Axial Computarizada*
TN. *True Negative*
TNR. *True Negative Rate*
TP. *True Positive*
TPR. *True Positive Rate*

1. INTRODUCCIÓN

1.1. Contexto

El cáncer es el nombre que se da a un conjunto de enfermedades relacionadas. En todos los tipos de cáncer, algunas de las células del cuerpo empiezan a dividirse sin detenerse y se diseminan a los tejidos del derredor [1].

Dentro de los distintos tipos de cáncer, *glioma* es el término utilizado para referirse a los tumores cerebrales primarios que, a su vez, se clasifican dependiendo de la célula que lo origina. Son los tumores más comunes del sistema nervioso central y agrupan casi el 80% de todos los tipos de tumores del cerebro. En este grupo se incluye el Glioblastoma multiforme (GBM), el tumor más maligno [2].

Los principales tipos de tratamiento actualmente consisten en una combinación entre una cirugía que tiene como objetivo la resección máxima del tumor, radioterapia y quimioterapia. A pesar de estos tratamientos y los avances realizados, siguen sin ser muy eficaces y los pacientes tienen una esperanza de vida de 16-18 meses y tan solo el 3-5% sobrevive los 5 años [3]. A pesar de tener una incidencia menor de 10 por 100000 personas, su mal pronóstico lo convierte en un problema crucial de salud pública. Puede darse a cualquier edad, pero tiene mayor incidencia en el rango de 55 a 60 años.

Una detección precoz del cáncer es de suma importancia. La detección y diagnóstico mediante técnicas de imagen avanzadas reduce la necesidad de una cirugía exploratoria para ello. Además, la posibilidad de detectar y tratar un *glioma* de bajo grado, antes de que progrese a un GBM, puede marcar la diferencia entre la vida y la muerte [4]. Así, se podrían tratar cuando todavía son pequeños antes de que se conviertan en malignos y aún pueden ser curables mediante cirugía mínimamente invasiva o incluso por métodos de radiación estereotáctica como la radiocirugía [5]. Cabe destacar que no siempre es posible detectar un *glioma* de bajo grado ya que a veces se generan de nuevo sin evidencias de un precursor de menor grado [6]. La fundación del GBM afirma que una detección precoz podría llegar a posibilitar un tratamiento sin cirugía. También expone que varios estudios han demostrado que la resección del GBM puede estimular cualquier célula cancerígena no eliminada y hacer que crezca hasta un 75% más rápido que antes de la intervención, dando aún más importancia a la detección [7].

Los datos que se extraen para la valoración como para el seguimiento de esta enfermedad se realizan de forma manual habitualmente, siendo evaluados por profesionales para la toma de decisiones.

El proceso de diagnóstico es una actividad compleja y colaborativa que pide de un razonamiento clínico y recogida de información con el objetivo de determinar el estado de salud del paciente. Es un proceso (Figura 1) que se extiende en el tiempo, siendo iterativo y mientras se da la recogida de información, el objetivo es reducir las dudas,

limitar las distintas posibilidades para el resultado y desarrollar un entendimiento más preciso y completo del problema de salud del paciente.

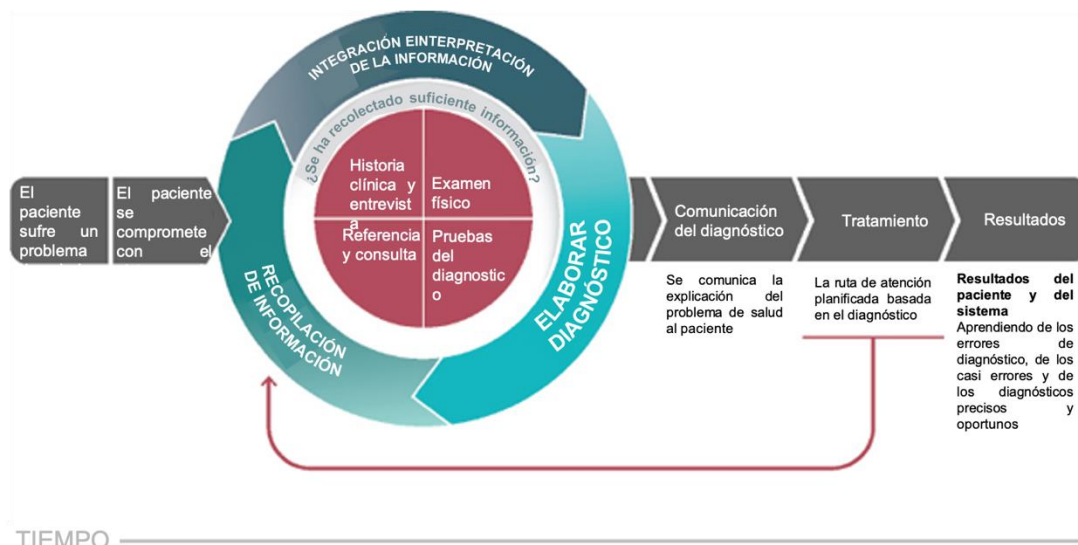


Figura 1. Proceso del diagnóstico. Figura traducida de [8].

La obtención de una correcta evaluación es clave, ya que proporciona información sobre el estado del paciente y es la información sobre la cual se basan los tratamientos posteriores para la mejora de este. Los servicios de salud cuentan con un error común para todos los ámbitos de estudio, los errores en el diagnóstico, que siguen produciéndose en una alta proporción.

Es probable que la mayoría de las personas experimente una valoración errónea en su vida, a veces con consecuencias devastadoras. Estos errores pueden causar daño a los pacientes al impedir o retrasar el tratamiento adecuado, proporcionar un tratamiento innecesario o perjudicial, o tener repercusiones psicológicas o financieras. Por ejemplo, se han encontrado los siguientes datos [8].

- Un cálculo conservador encontró que el 5% de los adultos de EE. UU. que buscan atención médica cada año experimentan un error de evaluación [9].
- La investigación durante décadas sobre las autopsias ha demostrado que estos errores contribuyen a aproximadamente el 10% de las muertes de pacientes.
- Las revisiones de los registros médicos sugieren que estas equivocaciones representan entre el 6 y el 17 % de los eventos adversos del hospital.
- Las confusiones son el principal tipo de reclamaciones por negligencia médica pagadas, tienen casi el doble de probabilidades de haber resultado en la muerte del paciente en comparación con otras reclamaciones, y representan la mayor proporción de los pagos totales.

A pesar de la presencia generalizada de los errores de diagnóstico y del riesgo de que se produzcan graves daños a los pacientes, estos no han sido muy considerados a la hora de mejorar la calidad y seguridad de los pacientes en los servicios de salud. Sin un enfoque dedicado a mejorar esto, los errores probablemente empeorarán a medida que

la prestación de la atención de la salud y el proceso de diagnóstico sigan aumentando en complejidad.

En un futuro, el objetivo es que la inteligencia artificial (IA) sea la encargada de realizar una primera valoración de una forma más precisa a la que pueda realizar el profesional, sirviendo de base para mejorar y formando parte de la experiencia de la medicina personalizada. Hoy en día existen sistemas de este tipo aplicadas a la oncología, como por ejemplo el sistema IBM Watson [10], encargado de asistir a los médicos a la hora de planificar el tratamiento mediante el análisis de datos clínicos, datos genéticos e imágenes médicas, realizando recomendaciones.

1.2. Objetivos

Teniendo todo esto en cuenta, se quiere desarrollar un sistema basado en IA que sea capaz de detectar un GBM partiendo de imágenes médicas adquiridas mediante resonancia magnética (RM) o tomografía axial computarizada (TAC). Asimismo, se quieren combinar las características del tumor extraídas de las imágenes como pueden ser su posición, tamaño, forma o textura con otros biomarcadores y hábitos de vida sobre los pacientes para tratar de encontrar alguna relación con la aparición del GBM.

El objetivo principal del proyecto es desarrollar un sistema inteligente que realice las siguientes funciones.

- Obtener un primer diagnóstico mediante la detección del tumor desde las imágenes médicas y extraer características de éste.
- Predecir la probabilidad de supervivencia del paciente.
- Establecer cuáles son los factores más determinantes para la supervivencia del paciente.

1.3. Planificación

Para la planificación del proyecto, este ha sido dividido en distintas fases que marcarán el desarrollo y son necesarias para alcanzar el objetivo fijado.

1. Elaboración de un estado del arte sobre la detección del GBM multiforme y de las técnicas empleadas para su diagnóstico. El estado del arte deberá contener información sobre los siguientes puntos.
 - Información sobre el impacto que tiene este juicio, un análisis de datos numéricos sobre la aparición del tumor, por ejemplo, sintomatología, tasa de mortalidad o tratamiento tras diagnóstico.
 - Información sobre sistemas de detección y evaluación que hayan sido desarrollados hasta la fecha y puedan ser encontrados en distintas publicaciones. Estos sistemas no deben ser necesariamente para el diagnóstico del GBM multiforme. Diferentes métricas y parámetros de evaluación del sistema inteligente que se encuentran en la literatura

- deberán de ser analizados para poder aplicar estos al desarrollo del sistema propio y realizar una evaluación correcta de este.
2. Obtención del conjunto de datos con los datos poblacionales y las imágenes médicas necesarias para el desarrollo del proyecto.
 - Selección de un conjunto de datos que contenga un mínimo de 50 sujetos, incluyendo imágenes y biomarcadores asociados a los sujetos.
 3. Análisis del problema y planteamiento de su resolución de forma clara. Definir los diferentes pasos por los que pasarán los datos para la obtención del sistema inteligente, nombrando todas las herramientas que vayan a usarse en el proyecto y el papel que tienen estas en la resolución de este.
 4. Aprendizaje del lenguaje de programación Python para su empleo en el desarrollo del proyecto.
 - Aprendizaje de Python básico.
 - Aprendizaje de las librerías de Python para el procesamiento de imagen, Machine Learning (ML) y Deep Learning (DL).
 5. Análisis de las imágenes médicas para la detección del GBM.
 - Procesamiento previo de las imágenes médicas para facilitar el paso posterior de identificación del tumor.
 - Localización del tumor mediante técnicas de procesamiento de imagen.
 - Extracción de marcadores a partir de la imagen procesada para el sistema de predicción.
 - Se plantea la detección de la zona del tumor mediante DL (identificación en la propia imagen de la zona del GBM) como mejora para el sistema. (Extra)
 6. Desarrollo del sistema de predicción mediante los datos poblacionales y los marcadores extraídos mediante ML / DL.
 - Selección de los parámetros poblacionales para el desarrollo del sistema.
 - Organización y procesamiento de los datos que van a ser empleados por el sistema inteligente.
 - Elaboración del sistema de detección y predicción del desarrollo del GBM en un futuro.
 7. Validación del sistema inteligente mediante el cálculo de distintas métricas de evaluación extraídas en el estado del arte. La fiabilidad del resultado final deberá ser evaluada de forma adecuada acorde a los métodos empleados en otras soluciones.

La siguiente fase se plantea como una fase opcional, la cual será desarrollada solamente en caso de contar con tiempo extra para el desarrollo del proyecto.

8. Desarrollo de una interfaz gráfica para la gestión de pacientes y exhibir los diagnósticos y las predicciones del sistema, además de datos estadísticos.
 - Definición de una base de datos (BD) para el almacenamiento de los datos de los pacientes y su gestión.
 - Interfaz gráfica que permita la gestión de los pacientes, permitiendo realizar las funciones CRUD (*Create, Read, Update and Delete*).

- Visualización de distintas variables estadísticas.

1.3.1. Diagrama de Gantt

El desarrollo del proyecto se va a realizar en un tiempo aproximado de 4 meses, dividido en bloques de dos meses. Aunque esto sea así, el desarrollo del proyecto también se plantea en paralelo al desarrollo de las clases. Por esta razón, se ha desarrollado un único diagrama de Gantt para la planificación, teniendo la posibilidad de poder adelantar o retrasar tareas en función del ritmo de trabajo.

Además de las fases técnicas definidas, cada uno de los bloques de dos meses tiene que ir acompañado de una memoria que recoja información sobre las tareas realizadas en ese periodo de tiempo y una presentación al finalizar el segundo bloque de tiempo dedicado al proyecto. Estos se definen como entregas o hitos, y su desarrollo deberá darse en paralelo al desarrollo del proyecto.

De esta forma, se expone el diagrama de Gantt desarrollado en la Figura 2.

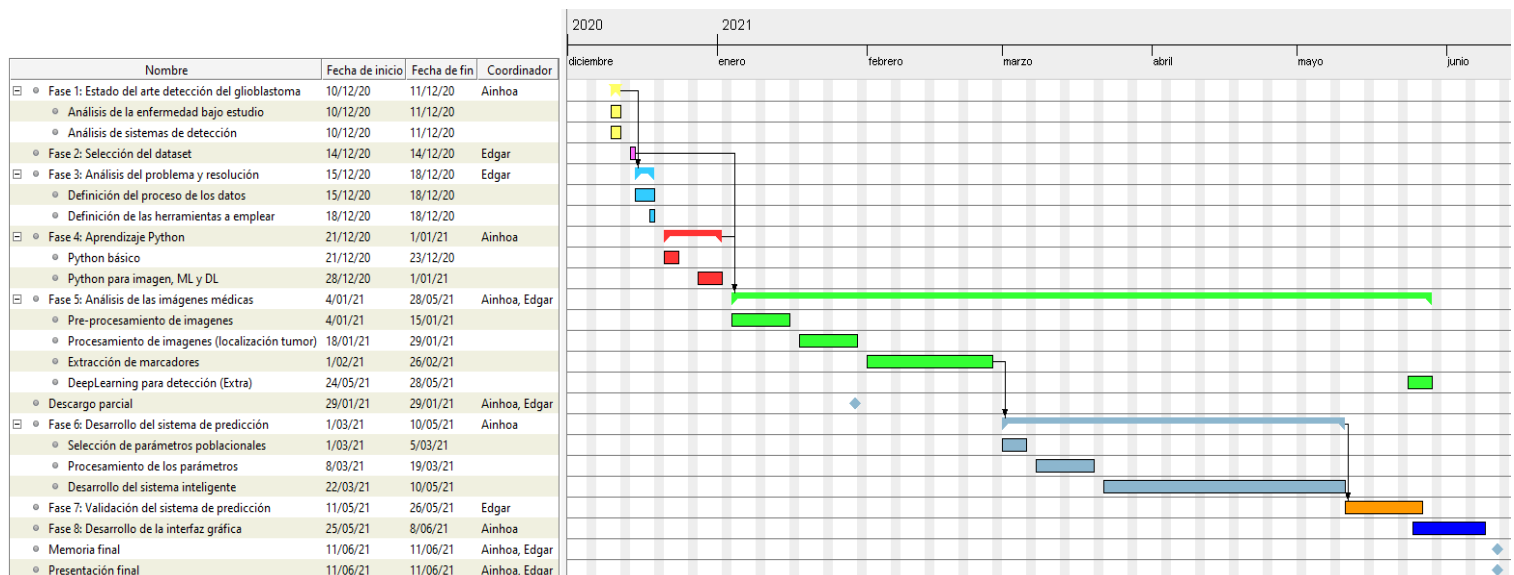


Figura 2: Diagrama de Gantt del proyecto.

2. ESTADO DEL ARTE

En esta sección se expone información sobre la enfermedad bajo estudio, el GBM multiforme, así como un resumen de ML y sus aplicaciones en el ámbito médico y para el procesamiento del *glioma*, información necesaria para la comprensión de la motivación de este proyecto.

2.1. Glioblastoma multiforme

Los siguientes apartados recogen información básica sobre el GBM, la adquisición de imágenes asociadas a estos y su análisis.

2.1.1. Clasificación de los *gliomas*

Más allá de la clasificación de los *gliomas* en diferentes tipos, estos tumores se clasifican de acuerdo con la presencia de características anaplásicas en tumores de bajo grado (o grado 2 de la Organización Mundial de la Salud (OMS)) y de alto grado (tanto grado 3 de la OMS o anaplásico y grado 4 de la OMS o GBM). Los *gliomas* de grado 1 son *gliomas* bien delimitados, y pueden ser curados si son completamente resecables. Los *gliomas* entre los grados 2 y 4 se encuentran entre los denominados *gliomas* difusos para los que no se dispone de un tratamiento curativo [11]. Esta clasificación se puede ver de una forma visual en la Figura 3.

Tipo \ Grado	OMS grado I	OMS grado II	OMS grado III	OMS grado IV
	Circunscripción	Bajo grado	Difuso	Alto grado
Astrocitoma	Astrocitoma pilocítico	Astrocitoma de grado bajo	Astrocitoma anaplásico	Glioblastoma
Oligodendroglioma		Oligodendroglioma de bajo grado	Oligodendroglioma anaplásico	
Oligoastrocitoma		Oligoastrocitoma de bajo grado	Oligoastrocitoma anaplásico	

Figura 3. Clasificación de los gliomas. Figura traducida de [11].

Los GBM pertenecen al grado 4 y son los más comunes (53%). Los *gliomas* difusos de grado 2 (difusos de bajo grado) y grado 3 (anaplásicos) representan aproximadamente el 30% de todos los *gliomas*. La edad media en el momento de la detección de estos *gliomas* es de 64, 43 y 56 años, respectivamente. Por otra parte, los *gliomas* pediátricos más comunes son los de grado 1 (predominantemente astrocitomas pilocíticos) que representan el 33% de los *gliomas* pediátricos [12].

2.1.2. Adquisición de imágenes

La adquisición de imágenes desempeña un papel central en el diagnóstico, la caracterización, la vigilancia y el seguimiento terapéutico de los tumores intracraneales.

La detección del GBM se basa tanto en técnicas invasivas como técnicas de imagen no invasivas.

El método por excelencia [13] para detectar los tumores son los escáneres de RM debido al contraste superior que ofrecen en tejidos blandos, lo que permite visualizar mejor la complejidad y heterogeneidad del tumor. En el caso de que el paciente no pueda someterse a una RM o los resultados sean contradictorios, se suele recomendar el uso de un TAC. La RM proporciona información estructural multiplanar de alta resolución y una caracterización de los tejidos sustancialmente mejorada en comparación con un TAC [14].

Existen varios tipos de imágenes de RM y cada una es más o menos adecuada a la hora de detectar distintas partes de los tumores. Entre estos tipos de RM, las imágenes ponderadas T1 con aumento de contraste (T1Gd con gadolinio, por ejemplo) son útiles para observar la alteración de la barrera hematoencefálica. El área necrótica puede ser apreciada por la parte hipo intensa del núcleo del tumor. En cambio, las imágenes ponderadas T2 y la *Fluid Attenuated Inversion Recovery* (FLAIR), se emplean para determinar la extensión del tumor y del edema [15]. Este comportamiento se aprecia en la Figura 4.

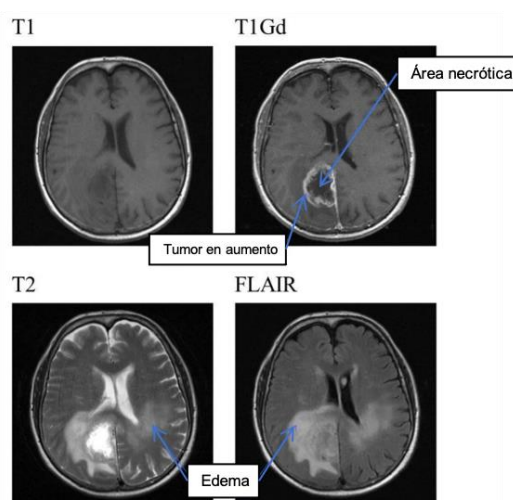


Figura 4. Tipos de RM para la detección de tumores. Figura traducida de [15].

2.1.3. Proceso de segmentación

La segmentación manual de los volúmenes tumorales requiere mucho tiempo y sufre de una gran variabilidad de un caso a otro. Esto limita su uso en la práctica clínica. Los métodos asistidos por ordenador reducen el tiempo de segmentación y demuestran una buena concordancia con segmentaciones manuales verificadas [16].

El diagrama de los principales pasos involucrados en la segmentación se puede ver en la Figura 5.

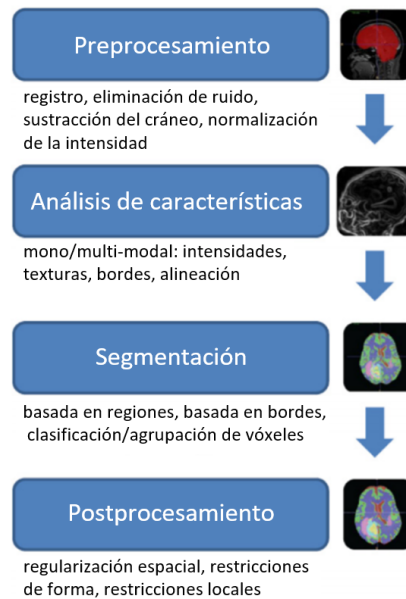


Figura 5. Bloques principales utilizados en la segmentación. Figura traducida de [17].

Los pasos de un proceso de segmentación se describen en los siguientes apartados.

a) Pre-procesamiento

Las imágenes suelen contener imperfecciones que pueden disminuir el rendimiento de la segmentación. Las principales causas de estas imperfecciones suelen ser una baja resolución, la presencia de artefactos, distorsión geométrica, un bajo contraste o un alto nivel de ruido. Mediante el pre-procesamiento de las imágenes se busca reducir lo máximo posible estos defectos para así maximizar el rendimiento del algoritmo [18].

El primer paso en el pre-procesamiento de imágenes suele centrarse en eliminar el ruido. Estas técnicas deben aplicarse cuidadosamente para no introducir nuevos artefactos que puedan hacer que la imagen se vuelva borrosa. Para ello, se pueden aplicar distintos filtros como pueden ser el filtro Gaussiano o el filtro de mediana entre otros [19].

Una de las técnicas recomendadas para mejorar las imágenes es la ecualización del histograma. Un histograma es una representación gráfica de la distribución de los tonos dentro de la figura. La ecualización del histograma consiste en cambiar la distribución de los valores convirtiéndola en uniforme. Así, se mejora el rango dinámico del histograma, aumentando de la misma forma su contraste [20].

La segmentación del cerebro, también denominado extracción del cerebro o eliminación del cráneo, es un paso importante de la mayoría de los estudios que emplean la RM como tecnología de imagen médica. Este paso se emplea por ejemplo en pre-procesamiento para el estudio volumétrico del cerebro [21], estudio de la esclerosis múltiple [22], evaluación de la esquizofrenia [23] o planificación de intervenciones neuroquirúrgicas [24].

Se trata de un paso de suma importancia para la posterior identificación de los distintos tejidos cerebrales, ya que determina una región de interés (ROI). De esta forma se reducen los errores de clasificación del tejido en la segmentación y a su vez se disminuye el tiempo de ejecución al eliminar los objetos de la imagen que no pertenecen al cerebro [25].

b) Análisis de características

La característica más común en la segmentación de tumores es el análisis de las intensidades presentes en la imagen. Esto se debe a que cada tejido presenta un tono distinto. También ocurre lo mismo con las texturas de las distintas partes del tumor. Los gradientes de intensidad o las características basadas en los bordes pueden utilizarse para la formación de un contorno del borde del tumor.

Dependiendo de los datos disponibles, todas estas características pueden ser computadas desde una sola modalidad de RM o de imágenes multimodales. Combinando las características de diferentes modalidades se logra mejorar los resultados de la segmentación [17].

c) Segmentación

La segmentación consiste en dividir las imágenes obtenidas en segmentos mediante la identificación de las distintas partes que la componen, facilitando así el estudio de las imágenes para su posterior análisis. Es un paso clave para el diagnóstico asistido por computadora, la cirugía, la radioterapia y la planificación de la quimioterapia. Los distintos tejidos pueden tener valores grises similares, lo que presenta desafíos para la segmentación precisa, repetible y estable de los *gliomas* [26].

Los distintos métodos existentes y sus características se explican más en detalle en la sección 2.1.4 Métodos de segmentación.

d) Post-procesamiento

Se puede incluir un paso de post-procesamiento que incluya operaciones morfológicas o de umbral. Un post-procesamiento basado en el ajuste de la región puede imponer la coherencia espacial después de la clasificación de los tejidos difusos [17].

2.1.4. Métodos de segmentación

El estándar actual se basa en mediciones ortogonales bidireccionales de un solo plano. Las principales investigaciones, así como los grandes ensayos clínicos se basan en esas medidas. Aun así, una volumetría tridimensional del tumor y del edema peritumoral ayudaría claramente a un mejor análisis del cambio longitudinal del tumor a lo largo del tiempo y a planificar el tratamiento. Los métodos de segmentación manual en 3D son tediosos, subjetivos y llevan mucho tiempo. Con los avances en las capacidades de computación y el refinamiento de los algoritmos, hay un creciente interés en los

métodos de segmentación de *gliomas* semiautomáticos y automatizados [27]. La Figura 6 muestra una segmentación semiautomática de un *glioma*.

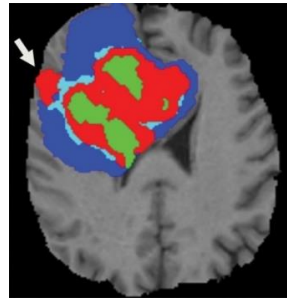


Figura 6. Segmentación de tumor cerebral semiautomática [27].

Se puede apreciar que la segmentación se divide en 4 secciones principales.

- Rojo: región en aumento.
- Verde: núcleo necrótico.
- Azul claro: tumor que no está en aumento.
- Azul oscuro: edema.

La diversidad de las aplicaciones del procesamiento de imágenes ha dado lugar al desarrollo de varias técnicas de segmentación de imágenes. Esto se debe a que no hay un único método adecuado para todas las imágenes. Los métodos de segmentación aplicados en RM pueden ser agrupados de la siguiente forma [28].

- Segmentación manual: un operador humano segmenta y etiqueta una imagen a mano. Es una tarea intensa y que requiere mucho tiempo. Aun así, todavía se sigue utilizando para definir un *ground truth* [29].
- Métodos basados en la intensidad: clasifica píxeles o vóxeles individuales basándose en su intensidad. En el caso de RM cerebrales, se distinguen 3 tipos de tejidos principales: materia blanca, materia gris y líquido cerebroespinal [30].
- Métodos basados en atlas: si se dispone de un atlas o plantilla del cerebro humano para una población de interés específica, este método puede ser una herramienta poderosa para la segmentación de RM cerebrales. El atlas contiene información sobre la anatomía del cerebro y se utiliza como referencia para segmentar nuevas imágenes. Su principal ventaja es la posibilidad de segmentar cualquier estructura cerebral disponible en el atlas sin ningún coste adicional [31]. Sin embargo, no es un método adecuado para la segmentación de tumores, dado que las variaciones en la localización, forma, heterogeneidad de los tejidos y el impacto en la topología del cerebro del paciente hace difícil la creación de un atlas adecuado [15].
- Métodos basados en la superficie: incluyen métodos basados en la superficie, como los modelos deformables que incluyen contornos y superficies activas. Los modelos deformables utilizan curvas o superficies paramétricas cerradas para delinear los límites de las regiones [32].
- Métodos de segmentación híbridos: dado que la selección de la técnica más apropiada para una aplicación determinada es una tarea difícil, puede ser

necesaria una combinación de varias técnicas para lograr el objetivo de la segmentación. La idea es combinar diferentes métodos de segmentación complementarios en un enfoque híbrido para evitar las desventajas de cada método por separado y mejorar la precisión de la segmentación [33].

2.2. Técnicas de análisis

2.2.1. Análisis de supervivencia

Para el análisis estadístico de las probabilidades de supervivencia de un paciente se emplean variables de tiempo hasta evento o *time to event variable*. De esta forma se analiza el tiempo en el que un paciente o participante sufre de un evento de interés. Este evento no tiene por qué tratarse de la muerte, dado que, al referirse a la supervivencia del participante, se busca que éste no sufra de dicho evento de interés.

Las técnicas que suponen una distribución normal para los datos no son útiles en estas aplicaciones debido a que el sesgo suele estar presente en la mayoría de sus aplicaciones. Por ejemplo, en un estudio que evalúa el tiempo hasta la recaída en pacientes de alto riesgo, la mayoría de los eventos pueden ocurrir al principio del seguimiento y muy pocos ocurren después.

Además, hay que sumarle la complicación adicional de no contar siempre con datos completos. En muchos estudios, los participantes se inscriben durante un periodo de tiempo (puede que incluso abandonen el estudio) y el estudio termina en una fecha concreta del calendario. Por lo tanto, a algunos participantes se les realiza un seguimiento más largo que a otros. Al no saber el verdadero tiempo de supervivencia se le denomina tener tiempos censurados [34].

a) Función de supervivencia

Para la estimación de la función o curva de supervivencia se emplea la información de estado del evento y tiempo de seguimiento. En esta curva se muestra el tiempo en el eje X y las probabilidades de supervivencia en el eje Y. Un ejemplo de esta curva se observa en la Figura 7, donde en el momento 0, todos los pacientes están vivos (probabilidad igual a 1) y la mediana muestra una supervivencia de 510 días. Es decir, el 50 % de los pacientes sobreviven a los 510 días.

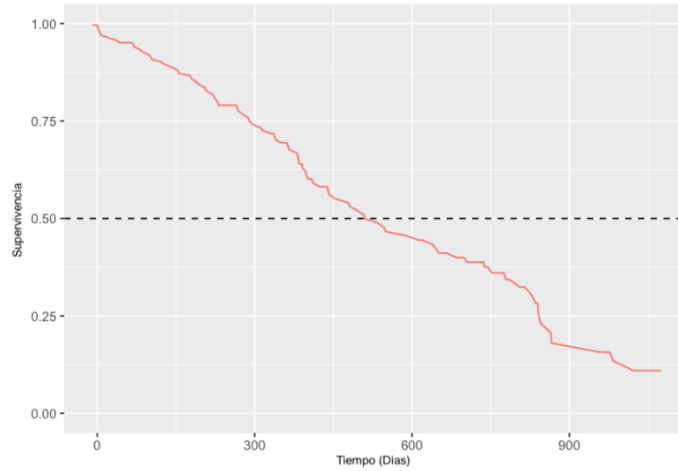


Figura 7. Ejemplo de curva de supervivencia.

Existen varias formas de estimar una función de supervivencia o una curva de supervivencia. Hay una serie de métodos paramétricos populares que se utilizan para modelar los datos de supervivencia, y difieren en términos de las suposiciones que se hacen sobre la distribución de los tiempos de supervivencia en la población. Algunas de las distribuciones más populares son las siguientes: exponencial, Weibull, Gompertz y log-normal [34].

Los métodos no paramétricos no hacen suposiciones sobre la variación la probabilidad de que una persona desarrolle el evento a lo largo del tiempo. Con los métodos no paramétricos se estima y traza la distribución de supervivencia o la curva de supervivencia. Las curvas de supervivencia suelen representarse como funciones escalonadas.

b) Regresión de riesgos proporcionales de Cox

La regresión de riesgos proporcionales de Cox se implementa para relacionar varios factores de riesgo o exposiciones, considerados simultáneamente, con el tiempo de supervivencia. En este modelo la medida del efecto es la tasa de riesgo, que es el riesgo o la probabilidad de sufrir el acontecimiento de interés. Esta tasa de riesgo se intenta modelar mediante el uso de regresiones con una parte empírica que depende del tiempo y una parte exponencial que depende de las otras variables. Su representación se define en la ecuación (1), donde $h_0(t)$ se denomina el riesgo base, $\beta_k, k = 1, \dots, p$ son los coeficientes de la regresión y $x_k, k = 1, \dots, p$ son los factores de pronóstico [35].

$$h(t) = h_0(t)^{\beta_1 x_1 + \dots + \beta_p x_p} \quad (1)$$

c) Evaluación

El método más común para la evaluación de modelos de supervivencia es mediante el uso del índice de concordancia o *C-index*, introducido por Harrell *et al.* [36]. Este índice expone la probabilidad de que un paciente seleccionado al azar que haya

experimentado un evento tenga una puntuación de riesgo más alta que un paciente que no haya experimentado el evento [37].

El cálculo del valor se da de la siguiente forma. Para el paciente i , nuestro modelo de riesgo asigna una puntuación de riesgo η_i y un tiempo T_i . En un modelo de supervivencia, los pacientes que han tenido tiempos más cortos hasta el evento deberían tener puntuaciones de riesgo más altas. Para todas las parejas posibles i y j , siendo $i \neq j$, la pareja será concordante en el caso donde $T_i < T_j$ y $\eta_i > \eta_j$, y discordante si $T_i > T_j$ y $\eta_i > \eta_j$. En general la concordancia se define como se representa en la ecuación (2).

$$C = \frac{\# \text{ parejas concordantes}}{\# \text{ parejas concordantes} + \# \text{ parejas no concordantes}} \quad (2)$$

De [38] se extrae la siguiente ecuación para el cálculo de la concordancia.

$$C = \frac{\sum_{i,j} I(T_i > T_j) \cdot I(\eta_j > \eta_i) \cdot \Delta_j}{\sum_{i,j} I(T_i > T_j) \cdot \Delta_j} \quad (3)$$

Donde Δ_j determina si la variable ha sido completamente observada ($\Delta_j = 1$) o ha sido censurada ($\Delta_j = 0$).

Este índice también puede definirse como el resumen del área bajo la curva ROC o *Receiver Operating Characteristic*, profundizado en el apartado Curva ROC.

2.2.2. Métodos de Machine Learning

El aumento del conocimiento y comprensión de las enfermedades ha conllevado al crecimiento de la información y datos asociados a estos, en parte gracias a los avances que ha habido en herramientas de análisis cuantitativo y cualitativo en cuanto a mediciones de parámetros fisiológicos se refiere. Al contar con un campo de datos tan grande, este se convierte en un escenario ideal para el uso de técnicas de ML. El potencial de herramientas de ML como plataforma para la obtención de información de numerosas fuentes en un sistema para la asistencia en la toma de decisiones está siendo descubierto [39].

Todo proceso de aprendizaje consiste en dos fases distintas que se definen a continuación [40].

1. Estimación de las dependencias desconocidas de un sistema a partir de un conjunto de datos determinado
2. Empleo de las dependencias estimadas para predecir los nuevos productos del sistema.

El objetivo principal de las técnicas de ML es producir un modelo que pueda utilizarse para realizar una clasificación, predicción, estimación o cualquier otra tarea similar. Los algoritmos de ML existentes se pueden clasificar en dos grupos: aprendizaje supervisado y aprendizaje no supervisado. Además, existe la posibilidad de aplicar

algoritmos de aprendizaje semi-supervisado, lo que combina elementos de los dos grupos.

a) Aprendizaje supervisado

El aprendizaje supervisado se refiere a técnicas donde un modelo se entrena con una serie de entradas o características asociadas a un resultado relevante, con el objetivo de que cuando el modelo esté entrenado, este sea capaz de predecir el resultado mediante la introducción de nuevos datos. Los resultados de las predicciones de los modelos del aprendizaje supervisado pueden ser discretos o continuos.

A veces, para que el algoritmo de ML sea capaz de comprender la información de las entradas, por ejemplo, en el reconocimiento de imágenes o procesamiento del lenguaje, las variables (píxeles o palabras en este caso) deben ser procesadas para extraer de estos una serie de características identificables que puedan ser representadas en una matriz numérica. En el caso de las imágenes, una característica puede ser el color del píxel y en el caso de las palabras, el número de veces que aparece en un texto.

Una vez que un conjunto de datos se ha organizado en características y resultados, se le puede aplicar un algoritmo de ML. El algoritmo se mejora iterativamente para reducir el error de predicción mediante una técnica de optimización [41].

Es importante tener en cuenta que al entrenar los algoritmos de ML es posible sobreajustar el algoritmo a un conjunto de datos específico, lo que da lugar a un modelo de predicción que no generaliza bien a los nuevos datos. Esto se denomina *overfitting* y el riesgo de que este se dé puede reducirse mediante el uso de bases de datos de mayor tamaño, reducción del número de características seleccionando solo las más importantes o gracias a las técnicas de división de los datos que se explican en el apartado d) Evaluación de los sistemas.

Se distinguen dos tipos de algoritmos dentro del aprendizaje supervisado.

Clasificación

Un modelo que produce categorías discretas (a veces denominadas clases) se denomina algoritmo de clasificación. Por ejemplo, mediante este tipo de algoritmos se puede determinar si un tumor es benigno o maligno.

En la práctica, los algoritmos de clasificación devuelven la probabilidad de una clase (entre 0 para imposible y 1 para definitiva). Típicamente, transformaríamos cualquier probabilidad superior a 0,50 en una clase de 1, pero este umbral puede ser alterado para mejorar el rendimiento del algoritmo según sea necesario.

Existen distintos tipos de algoritmos que entran dentro de este grupo, algunos se nombran a continuación [42] y sus rendimientos se comparan en la Tabla 1.

- *Decision Trees* o árboles de decisión: se emplean para organizar gráficamente la información sobre posibles opciones, consecuencias y valor final. Se utilizan para decidir entre varios cursos de acción. Crean una representación visual de los

diversos riesgos, recompensas y valores potenciales de cada opción. Uno de los algoritmos más habituales es el *Random Forest* (RF) o bosques de decisión, que se basa en crear distintos árboles de decisión con diferentes combinaciones de datos y elegir la mejor solución mediante un sistema de votos [43].

- Clasificación mediante Naive Bayes: método probabilístico para la clasificación multiplicando las probabilidades individuales de cada par de atributos-valor. Este simple algoritmo asume la independencia entre los atributos.
- Redes neuronales artificiales (Artificial Neural Networks o ANN): sistema diseñado para simular la forma en la que el ser humano analiza y procesa la información. Se construyen mediante capas de neuronas interconectadas entre ellas. Normalmente una única capa es suficiente para la clasificación de la mayoría de conjunto de datos.
- *Support Vector Machines* (SVM): es una función que toma el espacio de entrada de baja dimensión y lo transforma en un espacio de mayor dimensión, es decir, convierte un problema no separable en un problema separable. Es muy útil cuando se trata de problemas donde la separación a realizar no es lineal y se puede emplear también para problemas de regresión.

Tabla 1. Resultados de los algoritmos de aprendizaje supervisado [42].

	Decision Trees	Naive Bayes	Redes Neuronales	SVM
Accuracy (ACC)	**	*	***	****
Velocidad aprendizaje	***	*****	*	*
Velocidad de clasificación	***	***	****	****
Manejo de atributos irrelevantes	***	**	*	****
Tolerancia a atributos redundantes	**	*	**	***
Prevención de <i>overfitting</i>	**	***	*	**

Regresión

Un modelo que devuelve una predicción de un valor continuo se conoce como algoritmo de regresión. El uso del término regresión en el ML varía de su uso en la estadística, donde la regresión se utiliza a menudo para referirse tanto a los resultados binarios (es decir, la regresión logística) como a la salida continua (es decir, la regresión lineal). Por ejemplo, este tipo de modelos se pueden emplear para predecir la esperanza de vida de un individuo o la dosis tolerable de quimioterapia.

Los algoritmos de regresión más relevantes se describen a continuación [44], [45].

- Regresión lineal simple: método estadístico que permite resumir y estudiar las relaciones entre dos variables continuas (cuantitativas). Este algoritmo asume la existencia de una relación lineal entre las variables de entrada y una variable de salida. En este caso, la variable de salida puede calcularse a partir de una combinación lineal de las variables de entrada. Cuando hay una única variable de entrada, el método se denomina regresión lineal simple. Cuando hay múltiples variables de entrada, se denomina regresión lineal múltiple.

- **Regresión LASSO:** el objetivo de este algoritmo es obtener el subconjunto de variables de entrada que reduzca el error de predicción. El algoritmo funciona imponiendo una restricción a los parámetros del modelo que hace que haya variables que se excluyan del modelo. Se trata de un algoritmo de selección de variables.
- **Support Vector Regression (SVR):** trabaja de la misma forma que SVM, pero esta vez aplicado a problemas de regresión, pero con mínimas modificaciones. Aunque la idea es la misma, reducir el error, el algoritmo de SVR es más complicado al de SVM. La mayor diferencia es que mientras que el problema de clasificación tiene como resultado un número, a este se le añade un rango de tolerancia para convertirlo en el resultado para una regresión.
- **Decision Trees:** de la misma forma que en el algoritmo de clasificación, crean una serie de nodos, pero a diferencia de crear resultados de una misma clase (0 o 1 en caso de ser binario), crea valores continuos.

b) Aprendizaje no supervisado

El aprendizaje no supervisado se refiere a técnicas donde el modelo se entrena con una serie de entradas o características que no tienen asociadas un resultado o etiqueta de la clase a la que pertenecen. De esta forma, el objetivo del sistema es desarrollar y organizar los datos, buscando características comunes o grupos entre ellos, y cambiando en base a los conocimientos internos.

El aprendizaje no supervisado incluye varias técnicas de reducción de dimensiones. A veces se emplean métodos no supervisados junto con los métodos de aprendizaje supervisado, ya que, al comprimir la información de un conjunto de datos en menos características o dimensiones, se pueden evitar problemas como el alto costo de los cálculos [41].

En la Figura 8 se muestra visualmente una técnica de reducción de dimensiones. En esta figura, los datos originales se muestran a la izquierda representados por distintas formas y se introducen en un algoritmo de aprendizaje no supervisado que agrupa los datos. Las características que no tienen suficiente relevancia se excluyen, reduciendo así el número de características dentro del conjunto de datos y permitiendo la agrupación.

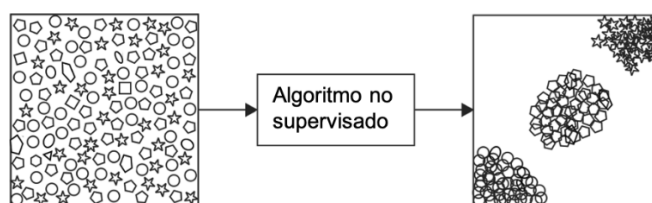


Figura 8. Ejemplo de aprendizaje no supervisado. Figura traducida de [41].

Los algoritmos de aprendizaje no supervisado pueden dividirse en dos grupos.

- **Clustering:** Un problema de *clustering* se presenta cuando se quiere descubrir las agrupaciones inherentes de los datos.

- *Asociación*: Un problema de aprendizaje de reglas de asociación es cuando se quiere descubrir reglas que describan grandes porciones de los datos. Por ejemplo, una regla de asociación puede ser que las personas que compran X también tienden a comprar Y.

Los algoritmos principales de aprendizaje no supervisado pertenecientes a estos dos grupos se presentan a continuación.

- *K-means clustering*: comienza con un primer grupo de centroides seleccionados al azar, que se utilizan como puntos de inicio para cada clúster, y luego realiza cálculos iterativos para optimizar las posiciones de los centroides, deteniéndose solo cuando los centroides se han estabilizado (agrupación exitosa) o se ha llegado al número máximo de iteraciones.
- *Principal Component Analysis* (PCA): técnica utilizada para reducir la dimensionalidad de los conjuntos de datos de alta dimensión, conservando la estructura original y las relaciones inherentes al conjunto de datos original.
- *Independent Component Analysis* (ICA): técnica para la obtención de bases linealmente independientes para los datos, es decir, identifica componentes o fuentes independientes de los datos.
- *Self-organizing Map* (SOM): aprende a categorizar los datos en base la distancia entre los clústeres dentro de ella. La técnica empleada se denominan mapas autoorganizados y se trata de mapas computacionales topográficos de inspiración biológica que aprenden mediante la auto-organización de sus neuronas [46].

c) Técnicas de Deep Learning

El DL es una forma de aprendizaje automático que permite a los ordenadores aprender de la experiencia y entender el mundo en términos de una jerarquía de conceptos. Dado que el ordenador reúne conocimientos a partir de la experiencia, no es necesario que un operador humano especifique formalmente todos los conocimientos que necesita el ordenador. La jerarquía de conceptos permite al ordenador aprender conceptos complicados construyéndolos a partir de otros más simples. Estas jerarquías tienen varias capas de profundidad [47]. Se refiere a ANN con varias capas.

Unas de las redes neurales profundas más populares son las redes neuronales convolucionales (CNN). Las CNN tienen un excelente rendimiento en problemas de ML, especialmente en las aplicaciones que tratan con datos de imágenes, visión por ordenador y en el procesamiento del lenguaje natural. El aspecto más beneficioso de las CNN es la reducción del número de parámetros en las ANN [48].

Las CNN se componen de tres tipos de capas principales: capas convolucionales, capas de agrupación y capas totalmente conectadas. La unión de estas capas forma una arquitectura de CNN como la de la Figura 9.

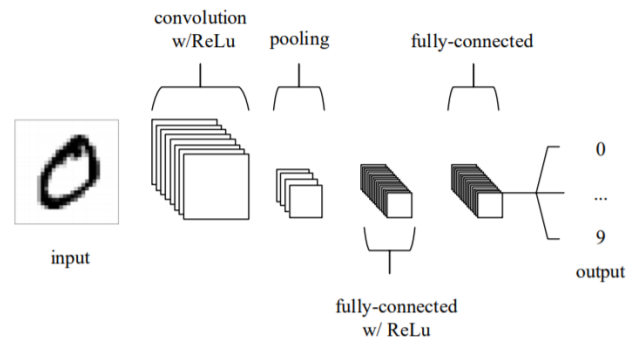


Figura 9: Arquitectura simple de una CNN, formada por 5 capas [49]

- La capa de entrada contiene los valores de los píxeles de la imagen.
- La capa convolucional determinará la salida de las neuronas que están conectadas a regiones locales de la entrada mediante el cálculo del producto escalar entre sus pesos y la región conectada al volumen de entrada. La unidad lineal rectificadora (ReLU) tiene como objetivo aplicar una función de activación “por elementos” a la salida de la activación producida por la capa anterior.
- La capa de agrupación realiza un *downsampling* en la dimensionalidad espacial de la entrada, reduciendo el número de parámetros dentro de esa activación.
- Las capas totalmente conectadas intentan producir puntuaciones de clase a partir de las activaciones, que se utilizan para la clasificación.

d) Evaluación de los sistemas

Los datos empleados para la fase de entrenamiento del modelo no pueden ser empleados para la evaluación de este, dado que el sistema recordará esa combinación concreta y realizará una predicción correcta para todos los datos. Para poder realizar la evaluación de los sistemas obtenidos de forma correcta, existen dos métodos de separación de datos distintos para la creación del modelo [50].

- *Holdout*: los datos de entrenamiento se subdividen en conjuntos de entrenamiento y de validación. El conjunto de entrenamiento se utiliza para entrenar la red, mientras que el conjunto de validación se utiliza para evaluar el rendimiento de la red mediante la estimación de su error medio absoluto [51]. Esta técnica se emplea dado su simplicidad, velocidad y flexibilidad, pero se asocia a una alta variabilidad, ya que las diferencias entre el conjunto de datos empleados para el entrenamiento y el testeo del modelo pueden afectar a la precisión de las predicciones.
- *Cross-validation*: divide los datos en un grupo para el entrenamiento y otro para la evaluación. La técnica más habitual se denomina *k-fold*, y consiste en dividir el conjunto de datos original en k grupos distintos del mismo tamaño, a los cuales se les denomina *folds*. El entrenamiento del modelo se repite k veces, empleando cada vez un conjunto para la validación y el resto para el entrenamiento. La estimación del error se promedia sobre todas las combinaciones para obtener la eficacia total de nuestro modelo.

Para poder evaluar correctamente el rendimiento del sistema se definen distintos parámetros. Por ejemplo, en el caso de la clasificación: True Positive (TP), True Negative (TN), False Positive (FP) y False Negative (FN) que se expresan de forma gráfica en la Figura 10, lo que se denomina matriz de confusión. Mediante estos parámetros se obtienen las distintas métricas que se explican a continuación.



Figura 10. Matriz de confusión. Figura traducida de [16].

Accuracy

Muestra el porcentaje de predicciones realizadas correctamente por el sistema. Se obtiene mediante la fórmula definida en la ecuación (4).

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

Sensitivity o Recall (True Positive Rate o TPR)

Se define como la capacidad de clasificar los elementos positivos como tal y se obtiene mediante la ecuación (5).

$$TPR = \frac{TP}{TP + FN} \quad (5)$$

Specificity (True Negative Rate o TNR)

Se define como la capacidad de clasificar los elementos negativos como tal y se obtiene mediante la ecuación (6).

$$TNR = \frac{TN}{TN + FP} \quad (6)$$

Precision o valor predictivo positivo (Positive Predictive Value o PPV)

Expresa cuales de los casos clasificados como positivos, cuales se corresponden en la realidad con un caso positivo. Se obtiene mediante la ecuación (7).

$$PPV = \frac{TP}{TP + FP} \quad (7)$$

Valor predictivo negativo (Negative Predictive Value o NPV)

Expresa cuales de entre los casos clasificados como negativos, cuales se corresponden en la realidad con un caso negativo. Se obtiene mediante la ecuación (8).

$$NPV = \frac{TN}{TN + FN} \quad (8)$$

Valor F1

Se trata de la media armónica entre el valor de *precision* y *recall*. De esta forma con una única métrica se evalúan ambos parámetros y de esta forma, cuanto mayor sea el valor de F1, mejor. Se calcula de la forma expresada en la ecuación (9).

$$F1 = \frac{2}{\frac{1}{PPV} + \frac{1}{TPR}} = \frac{2 * PPV * TPR}{PPV + TPR} \quad (9)$$

Uno de los problemas que presenta esta métrica es que *precision* y *recall* tienen el mismo peso, por lo que dependiendo de la aplicación habrá que buscar otras métricas para mejorar esta interpretación.

Curva PR

Se trata de la curva entre la *precision* y el *recall* para distintos umbrales de probabilidad, lo que permite el estudio del equilibrio entre la tasa de positivos reales y la tasa de negativos reales. Un ejemplo de esta curva se observa en la Figura 11. De esta forma, la parte superior izquierda de la gráfica es el espacio ideal donde ambos valores son los más altos posibles. A partir esta curva, también se calcula el *Area Under the Curve* (AUC), que sirve como resumen del rendimiento del modelo, siendo este mejor cuanto mayor sea el valor del AUC.

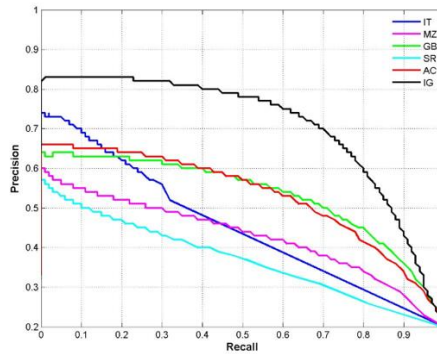


Figura 11. Curva PR para distintos valores de umbral [52].

Curva ROC

Herramienta útil para la evaluación de sistemas de clasificación binarios. La curva ROC representa el TPR o *recall* y el *False Positive Rate* (FPR), cuya formula está representada por la ecuación que se encuentra en (10).

$$FPR = 1 - TNR = \frac{FP}{TN + FP} \quad (10)$$

En la curva ROC, la parte superior izquierda de la gráfica representa una clasificación perfecta realizada por el sistema (Figura 12). De esta forma se compara el rendimiento del sistema con los resultados de un sistema donde la clasificación se realiza de forma aleatoria (representada mediante una línea roja en la Figura 12). A

partir esta curva también se calcula el AUC, que sirve como resumen del rendimiento del modelo, siendo este mejor cuanto mayor sea el valor del AUC.

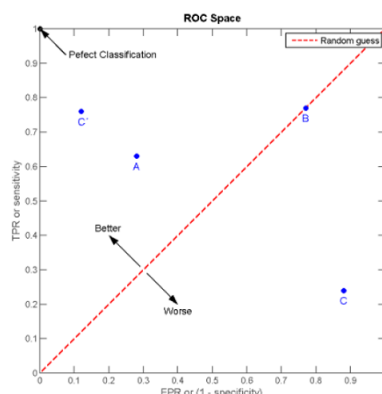


Figura 12. Ejemplo de curva ROC [52].

Evaluación de un modelo entrenado con datos descompensados

Trabajar con un dataset descompensado es un problema habitual a la hora de implementar un algoritmo de ML. Esto se da cuando, a la hora de crear un modelo para un problema de clasificación, el número de observaciones de una clase es mucho mayor a la del resto de clases. Un conjunto de datos de estas características puede llevar al modelo a devolver siempre la clase que más se repite. Esto se da en distintos campos de trabajo, entre ellos la medicina [53].

A la hora de evaluar los modelos, hay algunas métricas que no son adecuadas para detectar este tipo de comportamientos, como la métrica *accuracy*. Sin embargo, métricas como la matriz de confusión, precision, recall y F1, y la curva ROC, permiten detectar este tipo de comportamientos [54].

e) Aplicaciones

Dos de las áreas donde el uso de ML puede ser beneficioso dentro del campo de la medicina son el diagnóstico y predicción del resultado. Esto incluye la posibilidad de identificar algunas emergencias médicas como pueden ser las recaídas o las transiciones a enfermedades distintas.

Algunas de las aplicaciones de ML que se usan hoy en día en un contexto médico se expresan a continuación [55].

Chatbot

Programa informático con el cual se puede mantener una conversación gracias a la IA. Aplicados en este contexto, cuentan con capacidad para reconocer el lenguaje e identificar patrones en la expresión de los pacientes para realizar una evaluación potencial, prevenir enfermedades o recomendar un curso de acción apropiado.

Por ejemplo, una startup con sede en U.K. ha desarrollado la aplicación *Babylon Health* [56], que mediante el reconocimiento del lenguaje, compara los síntomas que

recibe de un usuario con una BD de enfermedades. En respuesta, recomendará un curso de acción apropiado basado en una combinación de los síntomas reportados, el historial del paciente y sus circunstancias. La aplicación también está diseñada para integrar los datos de los dispositivos del paciente para controlar los signos vitales como el ritmo cardíaco y el nivel de colesterol.

Patología

Se basa en el análisis de laboratorio de los fluidos corporales como la sangre y la orina, así como de los tejidos, para realizar diagnósticos de enfermedades. Se emplea la visión artificial y tecnologías de IA para realizar estas tareas que hasta este momento solo eran capaz de hacer patólogos mediante el uso de microscopios.

Con el objetivo mejorar la velocidad y la precisión de las valoraciones, un equipo de investigadores procedentes del Centro Médico Beth Israel Deaconess y de la Facultad de Medicina de Harvard han utilizado DL para entrenar un algoritmo capaz de integrar el reconocimiento de lenguaje y de imágenes para diagnosticar tumores en tejido. Este algoritmo demostró un 92% del éxito, algo menos que el 96% de acierto con el que cuentan los profesionales. Sin embargo, combinando el uso del algoritmo y el dictamen por parte de profesionales, el porcentaje del acierto obtenido asciende al 99,5% [57].

Enfermedades raras

Se emplea software de reconocimiento facial con técnicas de ML como ayuda a la diagnosis de enfermedades raras. Las fotos de los pacientes se analizan mediante el análisis facial y mediante DL se detectan fenotipos que se correlacionan con enfermedades genéticas raras.

La aplicación Face2Gene [58] es un software de reconocimiento facial que está combinando con el aprendizaje de la máquina para ayudar a los médicos a diagnosticar enfermedades raras (en este caso, de los rasgos dismórficos faciales). Las fotos de los pacientes se analizan mediante el análisis facial y mediante DL se detectan fenotipos que se correlacionan con enfermedades genéticas raras. La plataforma está actualmente disponible sólo para profesionales y así prevenir falsos positivos, soportando más de 7.500 trastornos.

Oncología

DL se está empleando para entrenar algoritmos para diversas funciones, como segmentar un tumor o realizar un seguimiento del desarrollo de un tumor, a un nivel comparable al de los profesionales médicos.

Investigadores de la Universidad de Stanford han desarrollado un algoritmo empleando DL, en concreto CNN para la diagnosis de cáncer de piel. Entrenan el algoritmo con un conjunto de datos de 129.450 imágenes clínicas, que se traducen en 2.032 enfermedades diferentes. Su desempeño se comprobó contra 21 dermatólogos certificados en imágenes clínicas probadas por biopsia, demostrando que el algoritmo desarrollado logra un rendimiento a la par de todos los expertos [59].

2.3. Detección de Glioblastoma mediante Machine Learning

Con la introducción de las nuevas tecnologías en la medicina, se han puesto grandes cantidades de datos sobre el cáncer a disposición de la investigación científica. Sin embargo, la predicción precisa de los resultados de una enfermedad sigue siendo un reto para los médicos, lo que ha provocado que métodos de ML se hayan convertido en una herramienta popular para investigadores en este campo de estudio. Estas técnicas pueden descubrir e identificar patrones y relaciones entre ellos a partir de complejos conjuntos de datos, que a simple vista pueden pasar desapercibidos por el análisis de un profesional. Además, algunas técnicas son capaces de predecir eficazmente los resultados futuros de un tipo de cáncer.

En los últimos años se han desarrollado nuevos enfoques de IA para *gliomas* con los que poder predecir su grado y genómica a partir de imágenes, automatizar su diagnóstico y proporcionar información sobre su pronóstico [60]. Técnicas nuevas de DL, así como redes neuronales y CNN han facilitado la extracción automatizada de datos clínicos relevantes para la planificación del tratamiento y la monitorización posterior al mismo [61].

2.3.1. Técnicas tradicionales de Machine Learning

En este apartado se presentan ejemplos de técnicas tradicionales de ML tanto para la extracción y predicción de distintos parámetros del GBM, así como para su segmentación.

a) Predicción

La radiómica, la extracción de alto rendimiento de características de imágenes cuantitativas tiene el potencial de proporcionar información sobre la fisiopatología subyacente y de desarrollar biomarcadores que son difíciles de descubrir mediante la inspección visual. Normalmente, la radiómica incluye las siguientes fases: pre-procesamiento de las imágenes, estimación de características, selección de características, clasificación y evaluación [12]. Varios estudios han demostrado que la combinación de la radiómica con técnicas tradicionales de ML es capaz de diferenciar el GBM de otros tipos de cáncer como un linfoma o tumores cerebrales metastásicos [62], [63], prediciendo información genética sobre el tumor, así como su posible respuesta ante el tratamiento o la posibilidad de sobrevivir a él [64].

Se han hecho estudios basados en características extraídas de imágenes de RM con las que, mediante un algoritmo de ML, poder predecir la expectativa de vida y el subtipo molecular del tumor. Utilizando la validación cruzada se asegura la generalización de los pronosticadores creados a nuevos pacientes. A pesar de las diversas correlaciones entre las características de las imágenes, la expresión genómica y la supervivencia reportadas en la literatura, ningún análisis ha sido lo suficientemente

contundente como para entrar en la práctica clínica. Por lo tanto, existe una dificultad considerable a la hora de determinar los biomarcadores que han de usarse y cuáles proporcionarán el resultado más exacto y reproducible. Por ejemplo, la Figura 13 muestra la distribución de las características más predictivas que marcan la supervivencia del paciente.

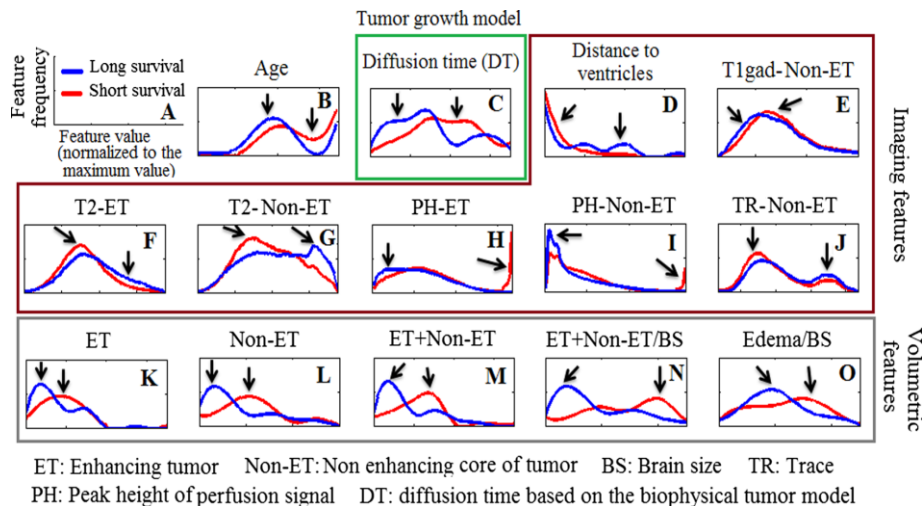


Figura 13. Histogramas de la distribución de las características más predictivas de la esperanza de vida [65].

Como demuestran los histogramas, es frecuente que existan pequeñas diferencias entre los supervivientes a corto y a largo plazo, y sólo surgen patrones distintivos que predicen la supervivencia mediante la integración de múltiples rasgos. Teniendo esto en cuenta, se han logrado predecir de forma no invasiva y precisa las variables clínicas y genómicas de los pacientes. La esperanza de vida de los pacientes (corta, media, larga) se ha predicho con un ACC del 80% mientras que en la clasificación del subtipo molecular se ha obtenido un 76% [65].

También se han realizado estudios para tratar de mejorar la clasificación de los tejidos de GBM heterogéneos. Así, se ha predicho el grado del *glioma* presente en los pacientes. Para ello, se han extraído varias características haciendo uso de radiómica para, después, seleccionar las más significativas. Tras esto, se ha entrenado y probado un método de clasificación usando árboles de decisión RF. Se ha obtenido una predicción consistente del grado del tumor en el 96% de los casos analizados [66]. Estudios realizados mediante regresión logística han logrado diferenciar entre *gliomas* de bajo grado y GBMs con un Accuracy del 93%, una sensibilidad del 97% y un AUC de 0.94 [67].

b) Segmentación

Las SVM son una herramienta útil para tareas de clasificación por lo que se han llegado a utilizar para la segmentación de tumores. Cabe destacar que asumen que los datos (vóxeles individuales) están distribuidos independiente e idénticamente lo cual no es muy adecuado para aplicaciones como la segmentación. La técnica de RF [68] o la de Campos Aleatorios de Markov (MRF) [69], son otras técnicas de clasificación populares que también se han aplicado en tareas de segmentación de tumores. Aun así, para superar las desventajas de estos métodos, otros modelos como *Support Vector Random*

Field (SVRF) han sido estudiados y se ha determinado que muestra un mejor resultado que las técnicas anteriormente citadas [70].

Se han realizado estudios para tratar de segmentar no solo el tumor sino realizar también la segmentación específica de tres tipos de tejidos relevantes de los *gliomas* de alto grado como las células activas, el núcleo necrótico y el edema. Para ello, se ha hecho uso de bosques de decisión como método de clasificación discriminatorio y, después, han sido combinados con un modelo generativo de la apariencia de los tejidos. Esto se logra utilizando estimaciones iniciales de la probabilidad de los tejidos basadas en modelos de mezcla Gaussianos entrenados como entradas adicionales para el bosque, junto con los datos de las RMs [71].

2.3.2. Nuevas técnicas de Deep Learning

En este apartado se exponen distintos métodos de DL enfocados en la extracción y predicción de características del GBM del mismo modo que se explica su aplicación en la segmentación.

a) Predicción

La extracción de características relevantes es una tarea crítica en la radiómica. Muchos estudios utilizan características extraídas a mano o explícitamente diseñadas. Estas incluyen la forma, intensidad y textura de los tumores entre otras. Aun así, estos rasgos de imagen son poco profundos y de bajo orden y es posible que no caractericen plenamente la heterogeneidad de la imagen, por lo que pueden limitar el potencial del modelo radiómico. El aprendizaje por transferencia puede incorporarse al modelo radiómico actual para la extracción de características profundas de varias capas de una CNN. El flujo del análisis radiómico se expone en la Figura 14.

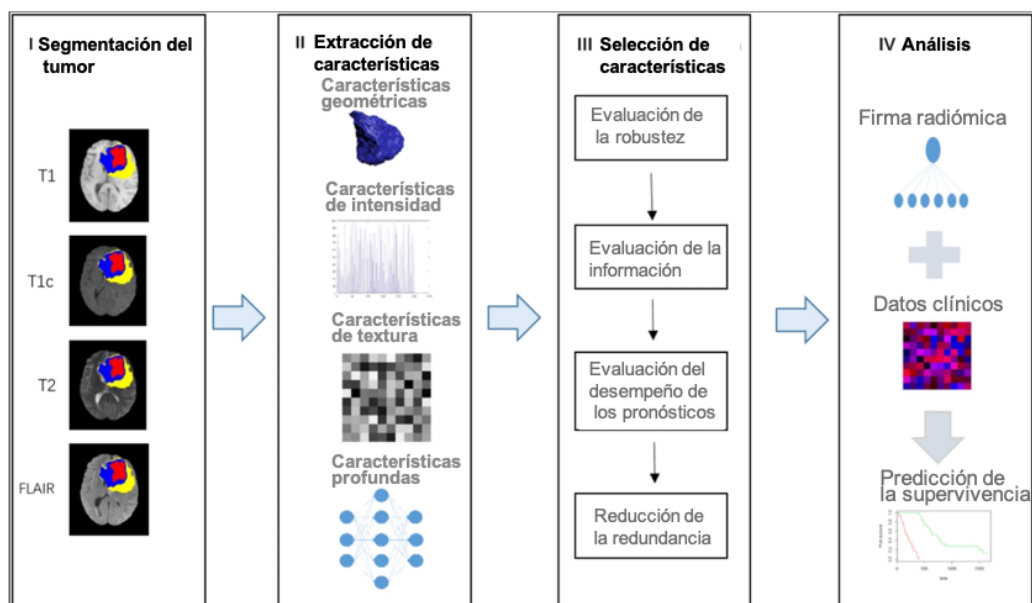


Figura 14. Flujo de trabajo del análisis radiómico [72].

Estos rasgos profundos contienen información más abstracta de las imágenes médicas y pueden proporcionar más patrones de predicción en comparación con los rasgos extraídos a mano [72].

b) Segmentación

Las CNN son el método asistido por ordenador puntero para la segmentación del GBM. Superan a los métodos alternativos que utilizan clasificadores de decisión independientes para extraer características de textura e intensidad. Se utiliza un conjunto de datos pre-annotados para entrenar la arquitectura de las CNN y realizar una serie de convoluciones matemáticas a través de múltiples capas interdependientes. Así se determina la relación entre las imágenes de entrada y las de salida. Después, la CNN puede ser validada en distintos conjuntos de datos de prueba identificando el fondo de la imagen de RM y las tres regiones jerárquicas del tumor [73].

Por ejemplo, DeepMedic es una CNN tridimensional con una alta calificación en el desafío de segmentación de tumores cerebrales (BRATS). DeepMedic asigna clases a cada vóxel de forma independiente usando la intensidad y la información de características locales. Ha demostrado que las mediciones volumétricas de las subregiones del tumor pueden medirse con gran precisión a partir de imágenes preoperatorias de RM. Además, la CNN puede integrarse en el flujo de trabajo radiológico para acortar el tiempo de segmentación hasta 20 minutos por paciente [16].

Otros estudios han evaluado la implementación de una segmentación completamente automática usando DL. Brevemente, la red de segmentación consiste de 3 instancias en cascada de una CNN bidimensional [74]. Esta arquitectura se puede observar en la Figura 15.

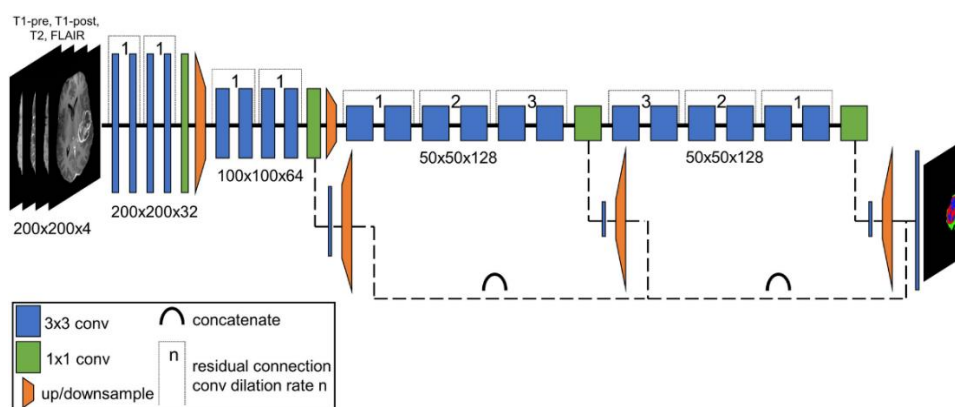


Figura 15. CNN usada para la segmentación de tumores cerebrales [74]

Además, también se ha seguido un enfoque basado en la radiómica para predecir el estado de varios biomarcadores genéticos comunes y clínicamente relevantes en los GBM utilizando únicamente imágenes preoperatorias de RM. El método desarrollado ha demostrado ser útil para predecir distintas mutaciones en los GBMs. Aun así, el rendimiento de estas técnicas de DL se puede ver afectado por el coste computacional que requieren sus arquitecturas complejas y la necesidad de grandes cantidades de datos de entrenamiento para obtener buenos resultados [27].

3. DISEÑO DE LA SOLUCIÓN

Con los datos que se han recogido hasta el momento en este documento, se ha realizado un diseño de la solución al problema definido para la realización de los objetivos marcados. Este diseño se describe en los siguientes apartados.

3.1. Requisitos técnicos

Para definir los servicios que va a realizar el sistema se han definido una serie de requisitos funcionales con los que debe cumplir.

- El sistema permitirá la visualización de las imágenes médicas y los biomarcadores de los pacientes almacenados en la BD.
- El sistema permitirá la introducción de nuevos pacientes (imágenes y datos poblacionales) y se guardarán los nuevos datos en la BD.
- El sistema realizará una detección del GBM a partir de las imágenes introducidas.
- Los resultados obtenidos de la detección deben ser mostrados, incluyendo las imágenes y datos necesarios para la comprensión del resultado obtenido de forma sencilla.
- Los resultados anteriores permitirán la confirmación del diagnóstico del paciente por parte del profesional médico correspondiente.
- Tras la confirmación de un profesional, el modelo será reentrenado usando los datos del paciente introducido.
- El sistema realizará una predicción del desarrollo del GBM a partir de imágenes y biomarcadores cuando se confirme el dictamen. Así, indicará si se prevé que el paciente sobreviva a los siguientes 12 meses o no.
- El sistema permitirá generar un informe en un documento de texto a partir de los datos del paciente.
- El sistema mostrará datos estadísticos que puedan ser obtenidos a partir de los datos de la BD.
- El sistema por desarrollar debe basarse en datos obtenidos de al menos 50 pacientes para asegurar la fiabilidad del sistema de predicción.
- La BD deberá contener datos del seguimiento del paciente de un mínimo de 12 meses para posibilitar la predicción de la supervivencia.
- El sistema deberá ser capaz de localizar el tumor con una precisión de $\pm 3-4\text{mm}$.
- El sistema deberá evaluarse empleando distintos parámetros y pruebas estadísticas. El sistema tiene que ser capaz de realizar las predicciones con una sensibilidad del 0.7, especificidad del 0.7, ACC del 70%, VPP y VPN del 70% y AUC del ROC del 0.8.
- Identificar las variables más relevantes que contribuyen a la predicción.

Como objetivo adicional, si se logran todos los puntos anteriores dentro de los límites marcados, los resultados obtenidos en la predicción y el diagnóstico, así como, en caso de tener un dictamen desfavorable, la imagen médica con el Glioblastoma

detectado, deben mostrarse al usuario en una interfaz gráfica. Esta interfaz también debe posibilitar la introducción de nuevos pacientes para ser evaluados. Además, dicha interfaz debe ser intuitiva y fácil de usar para todo tipo de usuarios.

Se han definido una serie de requisitos del software para la organización y desarrollo de distintos apartados del proyecto.

- El sistema será desarrollado en el lenguaje de programación Python, siguiendo el estándar PEP8 para el estilo del código.
- El código será analizado mediante SonarQube para el cálculo de distintas métricas que determinen la calidad del código y que se cumplen los estándares definidos en PEP8.
- Git se empleará como herramienta de control de versiones, almacenando el repositorio de forma remota mediante GitHub.

3.2. Arquitectura

En la Figura 16 se muestra de forma esquemática la arquitectura del diseño de la solución planteado para llevar a cabo el proyecto.

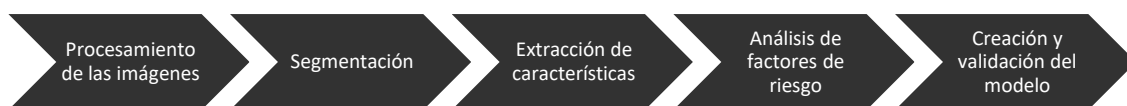


Figura 16. Esquema de la arquitectura del diseño de la solución.

En los siguientes puntos se hace una breve descripción de cada uno de los pasos, definiendo los puntos a realizar en cada paso.

- Pre-procesamiento de las imágenes. Las imágenes RM se han procesado para posibilitar la obtención de datos de los siguientes pasos. Las acciones realizadas engloban la mejora del contraste para la mejora de la relación señal/ruido o *Signal to Noise Ratio* (SNR). Además, se ha eliminado de las imágenes todos los elementos que no son el cerebro, por ejemplo, el cráneo.
- Segmentación. El objetivo de este paso es identificar y delimitar en la imagen el GBM y el edema presente a su alrededor.
- Extracción de características. A partir de los volúmenes segmentados se extraen distintas características de estos.
- Análisis de factores de riesgo. Con los datos poblacionales y las características obtenidas de los volúmenes del tumor y edema el objetivo es analizar cuales son los factores más determinantes para la supervivencia del paciente.
- Creación y validación del modelo. Se genera un modelo de ML que permita la predicción de la supervivencia del paciente.

3.3. Dataset

El conjunto de datos seleccionado pertenece al programa *Clinical Proteomic Tumor Analysis Consortium* (CPTAC) llevado a cabo por la *Office of Cancer Clinical Proteomics Research* (OCCPR) en el Instituto Nacional de Cáncer (NCI) de EE. UU. Este programa tiene el objetivo de impulsar la investigación del cáncer proporcionando datos y ensayos a investigadores de forma libre [75].

Este dataset recibe el nombre CPTAC-GBM [76] y se ha obtenido desde la página *The Cancer Imaging Archive*, formada por el NCI. Se trata de una plataforma de información de libre acceso y de código abierto para apoyar las iniciativas de búsqueda, desarrollo y educación que utilizan la imagen médica avanzada del cáncer [77].

Cuenta con dos tipos de imágenes médicas disponibles. Por un lado, se encuentran las imágenes de radiología de 66 pacientes, obtenidas por distintos métodos (RM, TAC y Radiografía computarizada o CR). De estas imágenes, se han seleccionado las obtenidas por RM. Por otro lado, se cuenta con imágenes patológicas (tejidos extraídos mediante biopsias) de 189 pacientes. Además, se incluyen datos clínicos de todos los pacientes en un archivo de formato JSON. A continuación, se presenta un ejemplo de los datos poblacionales disponibles de cada paciente. Para el desarrollo del proyecto se va a hacer uso de las imágenes de radiología además de los datos clínicos disponibles.

3.3.1. Pre-procesamiento del dataset

Se ha realizado un filtrado de estas con el fin de eliminar la información innecesaria y poder ver realmente los datos que se disponen. Esto se debe a que a cada paciente se le han realizado, en distintas fechas, varios escáneres con distintos tipos de RM con las que poder identificar distintas características, pero no se ha seguido un único proceso en todos los casos. Para el proyecto, dado que las imágenes se van a emplear para el diagnóstico, se han seleccionado las RMs que pertenecen al primer escáner realizado en el tiempo.

Teniendo esto en cuenta, se ha decidido identificar y mantener solamente las imágenes de RM axiales, coronales y sagitales de T2, T1, FLAIR y T1 con contraste. Se han analizado las imágenes de los 66 pacientes y se ha formado una tabla indicando la presencia o la ausencia de cada uno de estos tipos de imágenes en cada caso. Hace falta destacar que ciertos pacientes no contenían ninguna de estas imágenes o eran defectuosas haciendo imposible su tratamiento. Así, se han obtenido los datos de la Tabla 2.

Tabla 2. Información sobre el dataset

TOTAL	VÁLIDOS	T2 (%)	T1 (%)	FLAIR (%)	T1C (%)
66	56	75	75	62.5	100

Se puede apreciar que de los 66 pacientes iniciales solo 56 contienen información relevante. Las imágenes T1 con contraste son las más comunes por lo que los resultados

alcanzados partiendo de estas imágenes serán más contundentes que los obtenidos por ejemplo mediante las imágenes FLAIR, ya que son las que menos pacientes presentan.

3.4. Herramientas

Durante el desarrollo del proyecto, se ha hecho uso de distintas herramientas para dar respuesta a las partes que lo componen.

Por una parte, se ha decidido desarrollar el código del procesado de las RM en Python. Python es un lenguaje de programación interpretado, interactivo y orientado a objetos. Combina una notable potencia con una sintaxis clara [78]. Para este proyecto, se ha hecho uso de una versión 3.8 de Python.

Por otra parte, se ha optado por Visual Studio Code para la programación. Visual Studio Code es un editor de código simplificado que admite operaciones de desarrollo como la depuración, la ejecución de tareas y el control de versiones. Su objetivo es proporcionar las herramientas que un desarrollador necesita para un ciclo rápido de construcción y depuración de código [79]. Mediante una extensión es posible obtener soporte de Python incluyendo nuevas características como el *linting* que destaca los problemas sintácticos y estilísticos del código [80]. Cabe destacar que se ha seguido la guía de estilo PEP8 como base para el *linting* [81]. Además, se han usado las siguientes librerías principales.

- numpy 1.18.5 para la computación numérica y manejo de arrays.
- pydicom 2.1.2 para la lectura y escritura de archivos médicos DICOM.
- scikit-image 0.18.1 para el procesamiento de imágenes.
- medpy 0.4.0 para el procesamiento y filtrado de imágenes.
- scipy 1.6.0 para operaciones morfológicas y filtrado de imágenes.
- matplotlib 3.2.2 para la visualización de figuras.
- scikit-learn 0.24.1 para módulos de ML y minado de datos.

Para el desarrollo del análisis de datos se ha empleado el lenguaje de programación R, lenguaje y entorno para la computación estadística y creación de gráficos [82]. El editor de código RStudio se ha empleado para la programación. Se han empleado las siguientes librerías principales.

- jsonlite 1.7.2 para generar y leer archivos JSON.
- corrplot 0.84 para visualizar matrices de correlación.
- Amelia 1.7.6 para el análisis de valores desconocidos.
- caret 6.0-86 para el entrenamiento de clasificación y regresión de ML.
- e1071 1.7-4 para funciones de estadística.
- survival 3.2-11 para el análisis de supervivencia.
- ranger 0.12.1 para el uso del algoritmo RF en el análisis de supervivencia.
- mice 3.13.0 para la imputación de variables.
- ggplot2 3.3.3 para la visualización de gráficas.

4. DESARROLLO

En el siguiente apartado se describen los pasos ejecutados para la resolución del problema, siguiendo las acciones definidas en el apartado 3 DISEÑO DE LA SOLUCIÓN.

4.1. Pre-procesamiento de imágenes

Antes de realizar la segmentación del *glioma* dentro de la RM, se han llevado a cabo dos pasos para mejorar la calidad de la segmentación y la extracción del *glioma*. Para obtener esto, se ha mejorado la imagen y se ha eliminado todo tejido que no forma parte del cerebro de la imagen.

4.1.1. Estandarización de las escalas de intensidades

La falta de estándares en escalas de intensidades de las RM provoca muchas dificultades a la hora de mostrar y analizar las imágenes. Con el objetivo de mejorar esto se ha empleado un método de dos pasos de forma que, para las RM de un mismo protocolo y región del cuerpo, intensidades similares se correspondan a tejidos similares [83].

En el primer paso, partiendo de los volúmenes de un mismo protocolo de RM, se estiman los parámetros o *landmarks* de un histograma estándar. Esto solo se ejecuta una única vez. En el segundo paso, se transforma cada volumen para que los parámetros de su histograma se correspondan con aquellos del histograma estándar. Los parámetros se muestran en la Figura 17, explicando los significados de cada variable a continuación.

- m_1 y m_2 se corresponden a las intensidades máximas y mínimas presentes en el histograma
- μ es el segundo modo del histograma.
- p_1 y p_2 se corresponden a los percentiles máximos y mínimos de intensidad. [84] determina que los valores óptimos para estos parámetros son de 0 para p_1 y 99.8 para p_2 .

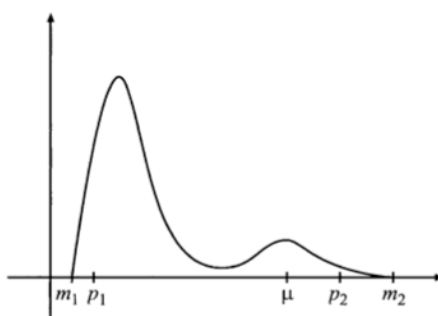


Figura 17. Parámetros correspondientes al histograma de una RM [83].

La Figura 18 muestra la comparación entre las RM originales y las obtenidas tras el proceso de estandarización llevado a cabo.

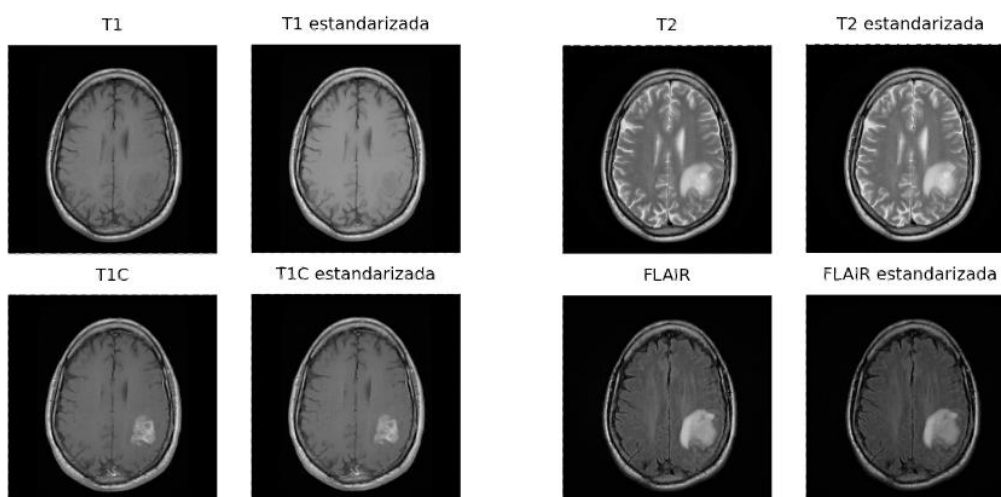


Figura 18. Comparación entre las RM originales y estandarizadas.

4.1.2. Eliminación del cráneo

Para la eliminación del cráneo se ha implementado un algoritmo basado en intensidades y bordes, que identifica el volumen correspondiente al cerebro dentro de la imagen y la extrae. Para ello, extrae una máscara que representa el volumen del cerebro.

El algoritmo implementado recibe el nombre de McStrip (*Minneapolis Consensus Strip*) [85], donde los autores definen una estrategia jerárquica para crear la máscara que defina el tejido que pertenece al cerebro. Se identifican 3 niveles distintos, y en cada nivel se define una máscara con un algoritmo distinto. Se da por hecho que cada una de estas máscaras va a generar un error a la hora de realizar la extracción, por lo que proponen llegar a un “acuerdo” entre las máscaras. Esto se obtiene mediante una regla basada en votos, donde para que un píxel sea aceptado como válido, debe tener al menos dos votos. La capacidad de obtener una extracción correcta dependerá de las influencias entre las máscaras, las cuales se definen en la Figura 19. El objetivo de este algoritmo es que la máscara obtenida al final sea más exacta que las intermedias, por lo que los diferentes algoritmos aplicados tienden a realizar distintos errores.

a) Máscara de Nivel 1

A la máscara de Nivel 1 le denominan *coarse mask* o máscara gruesa, ya que su único objetivo es el de restringir el campo de visión, es decir, delimitar dentro del volumen de la imagen médica lo que se corresponde con la cabeza del sujeto.

Con este objetivo se emplea el método de Otsu, técnica de *thresholding* que permite diferenciar el fondo de la imagen mediante el cálculo del umbral que marque la diferencia entre el fondo y los objetos sobre este. Este método prueba con distintos valores del umbral, y en aquel valor que consiga minimizar la varianza dentro del segmento es seleccionado [86].

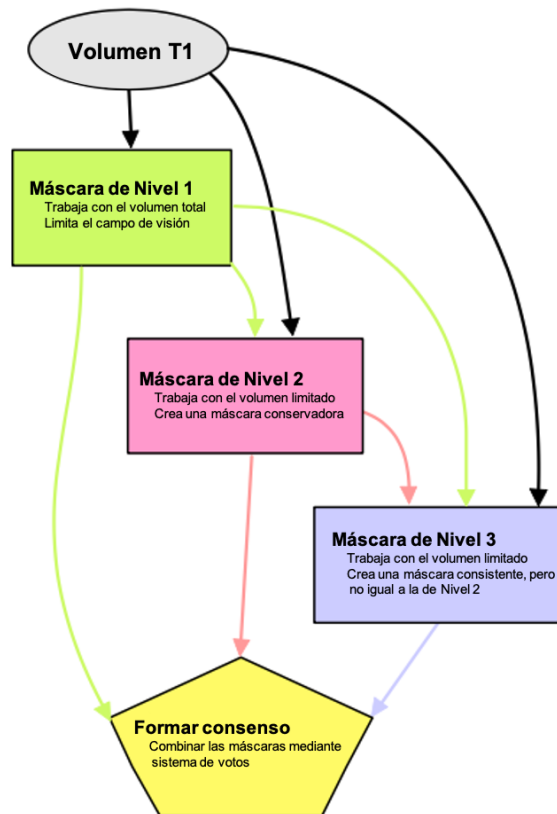


Figura 19. Algoritmo McStrip. Imagen traducida de [85].

En la Figura 20 se muestra (de izquierda a derecha) la imagen original, la máscara de Nivel 1 obtenida a partir de esa imagen y la máscara aplicada a la imagen original.

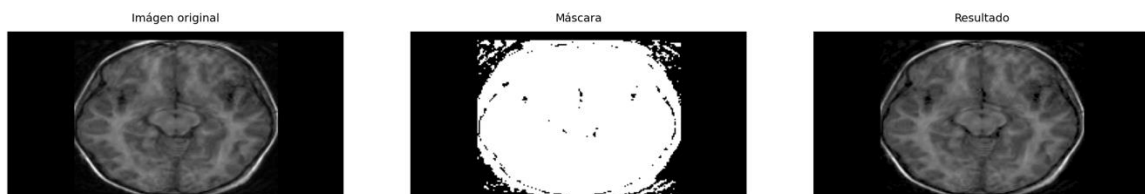


Figura 20. Máscara del Nivel 1 obtenida en el algoritmo McStrip.

a) Máscara de Nivel 2

La máscara de Nivel 2 se obtiene al aplicar el método de Otsu para la obtención de distintos valores de umbral a partir del volumen original. Mediante esto, en vez de diferenciar el fondo del objeto, se definen 5 clases distintas (fondo, líquido cerebroespinal, materia gris, materia blanca y cráneo) y se obtienen los valores de umbral que separan el histograma en estas clases.

Una vez se obtienen los valores de umbral, se aplica la máscara de nivel 1 al volumen y se eliminan los valores que corresponden al cráneo y al fondo. Se aplican operaciones morfológicas y se selecciona la región con el volumen más grande, que se corresponderá con el cerebro.

En la Figura 21 se observa (de izquierda a derecha) la imagen original, la máscara de Nivel 3 obtenida a partir de esa imagen y la máscara aplicada a la imagen original.



Figura 21. Máscara del Nivel 2 obtenida en el algoritmo McStrip.

b) Máscara de Nivel 3

El algoritmo implementado para la máscara de Nivel 3 se corresponde con el algoritmo que se denomina *Brain Surface Extractor* (BSE). La máscara del Nivel 1 se emplea como delimitante en este algoritmo, ya que se le aplica a la RM que se introduce.

El diagrama paso a paso del proceso de la extracción del cerebro para este algoritmo define de la Figura 22.

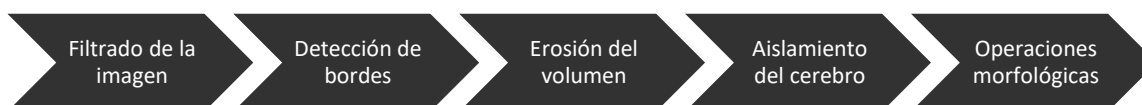


Figura 22. Diagrama del algoritmo BSE.

A continuación, se describe cada uno de los pasos ejecutados [87].

- Para mejorar la detección de bordes y eliminar irregularidades, se aplica un filtro espacial. Este filtro se denomina filtro anisotrópico de difusión, que suaviza las pequeñas diferencias de contraste e intensifica las grandes, lo que permite que los bordes de alto contraste se mantengan en la imagen mientras que las de bajo contraste se difuminan.
- Con el objetivo de detectar los bordes, se aplica el filtro *Laplacian of Gaussian* (LoG). Consiste en filtrar la imagen con un filtro gaussiano para posteriormente aplicarle un filtro Laplaciano. De esta forma, la imagen pasa a ser binaria conservando los valores mayores a 0 objeto y eliminando los menores.
- El volumen se erosiona para eliminar los pequeños objetos que pueden quedar adheridos al cerebro. El cerebro se aísla identificando el volumen más grande y eliminando el resto de las regiones. El volumen se dilata para compensar la erosión anterior. También se realizan operaciones morfológicas para eliminar posibles agujeros que queden en la máscara.

En la Figura 23 se observa (de izquierda a derecha) la imagen original, la máscara de Nivel 3 obtenida a partir de esa imagen y la máscara aplicada a la imagen original.



Figura 23. Máscara del Nivel 3 obtenida en el algoritmo McStrip.

c) Máscara de Consenso

La máscara de consenso se obtiene mediante una regla basada en votos. Partiendo de las máscaras de los tres niveles obtenidos anteriormente, para que un píxel sea aceptado, deberá estar incluido en al menos dos de las máscaras obtenidas. Después, se realizan operaciones morfológicas para llenar agujeros que puede haber presente en la máscara final.

En la Figura 24 se observa (de izquierda a derecha) la imagen original, la máscara de Nivel 3 obtenida a partir de esa imagen y la máscara aplicada a la imagen original.



Figura 24. Máscara de Consenso obtenida en el algoritmo McStrip.

4.2. Segmentación del *glioma* y del edema

Tras haber eliminado el cráneo de la imagen, se quiere aislar el tumor del resto del cerebro. Los tumores de mayor tamaño no suelen ser un gran problema a la hora de segmentarlos, pero las intensidades, el tamaño y la forma de otros tumores más pequeños se asemejan a otras partes sanas del cerebro lo que dificulta su detección. Para hacer frente a esto, se ha desarrollado un método de segmentación automático.

El algoritmo se ha desarrollado basándose en varias técnicas propuestas para la segmentación de tumores [88] y modificándolas para la aplicación en cuestión. La Figura 25 muestra el diagrama general de los distintos pasos implicados en el proceso.



Figura 25: Algoritmo implementado para la segmentación del tumor.

El punto de partida de este algoritmo son las imágenes obtenidas tras la eliminación del cráneo. El algoritmo desarrollado ha sido empleado tanto para segmentar el tumor en sí como para extraer el edema de las imágenes. Se han empleado distintos tipos de escáneres para cada caso partiendo de la teoría de la Figura 4: T1 y T1 con contraste para el área necrótica del tumor, y FLAIR o T2 para extraer el edema.

4.2.1. Mezcla de Gaussianas

Un modelo de mezcla Gaussianas es un modelo probabilístico que supone que todos los puntos de datos se generan a partir de una mezcla de un número finito de distribuciones Gaussianas con parámetros desconocidos [89].

En este caso, este método se aplica usando tres distribuciones Gaussianas. Del resultado de estas tres distribuciones, una corresponde al fondo negro de la imagen, otra representa las partes del cerebro de gran intensidad y la restante corresponde al resto del cerebro. Cada una de estas distribuciones obtenidas estará definida por su media y su covarianza. Estos datos se aplican como parámetros en las siguientes operaciones. Las distribuciones mencionadas se pueden apreciar en la Figura 26.

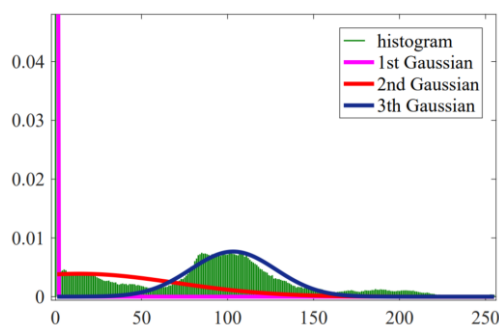


Figura 26: Aproximación del histograma mediante la mezcla de Gaussianas [88]

4.2.2. Reconstrucción morfológica de la escala de grises

Debido a que la zona del tumor tiende a ser de mayor intensidad que el resto del cerebro, se ha optado por la reconstrucción morfológica para identificar estos picos y facilitar la segmentación. La reconstrucción utiliza dos imágenes:

- Una semilla que especifica los valores que se propagan obtenida restando la media de la distribución Gaussianana de mayor covarianza a la imagen original.
- Una máscara que da el valor máximo permitido en cada píxel. La máscara es la imagen original de la que se dispone.

En este caso se ha aplicado la reconstrucción por dilatación, donde se conectan las regiones marcadas por los máximos locales en la semilla. Los máximos locales con valores superiores a la imagen semilla se truncan hasta el valor de la semilla mientras que los píxeles vecinos menores o iguales a esa semilla se conectan a la región saturada.

La imagen reconstruida resultante es exactamente igual a la imagen original, pero con los picos recortados. Por lo tanto, la sustracción de esta imagen reconstruida de la imagen original deja sólo sus picos de mayor intensidad. Esta operación se conoce como *h-dome* y deja los rasgos de altura '*h*' en la imagen sustraída. La Figura 27 ilustra el proceso descrito.

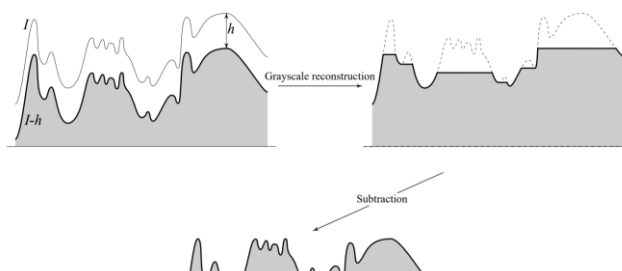


Figura 27: Determinación de los h-domes de una imagen en escala de grises [90]

4.2.3. *Threshold* binario adaptativo

A continuación, se aplica un *threshold* binario que solo mantendrá las zonas de más intensidad eliminando las que queden por debajo del umbral. Para determinar el umbral, se utiliza el método propuesto por Yen basado en la consideración de dos factores. La discrepancia entre la imagen umbralizada y la original, y el número de bits necesarios para representar la imagen umbralizada. Se hace uso de una función de coste que tiene en cuenta ambos factores para determinar automáticamente el umbral [91]. Así, se logra una máscara binaria con las regiones candidatas a contener el tumor.

4.2.4. Detección de la región con mayor intensidad media

Antes de empezar la selección de una región, se aplica una combinación de distintas operaciones morfológicas como la dilatación y la erosión para preparar la máscara binaria.

Debido a que la máscara creada puede estar compuesta por varias regiones, se ha desarrollado un algoritmo para detectar la región con mayor intensidad media. También se ha establecido un área mínima que debe superar cada región para evitar detectar un punto aislado sin mayor importancia. Adicionalmente, puede que el tumor se encuentre localizado en dos componentes colindantes, pero sin conexión entre ellos. Por ello, también se seleccionan las regiones que se encuentren dentro de un rango de distancia desde el componente principal seleccionado y su intensidad media sea mayor que el 80% de este componente principal.

4.2.5. Operaciones morfológicas

Finalmente, se realizan ciertas operaciones morfológicas como la dilatación, el llenado de agujeros, el cerrado o la erosión con el fin de cerrar posibles agujeros en la zona tumoral detectada y mejorar su definición. Esta máscara final obtenida se puede multiplicar con la imagen original para aislar el tumor de esta. El resultado obtenido en este punto, así como la evolución de la imagen en los pasos anteriores se muestra en la Figura 28.

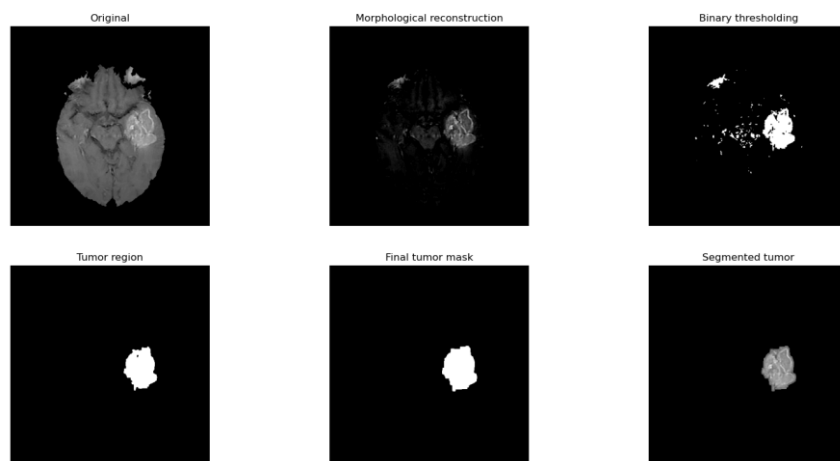


Figura 28: Proceso de segmentación.

4.2.6. Validación de la segmentación

Un tumor puede estar localizado en una o varias imágenes de un mismo escáner, pero no en todas a la vez. El algoritmo desarrollado no sabe de antemano en que imágenes se presenta una anomalía y en cuales no y puede haber segmentado una región de tejido normal del cerebro en ciertos casos. Por ello, es necesario seleccionar las imágenes que realmente contienen información de interés. Con este fin se ha desarrollado un algoritmo formado por dos partes: un primer filtrado que tienen que pasar todas las máscaras y una segunda parte donde se analiza la continuidad de estas.

a) Filtrado de máscaras

Se analizan distintas medidas para cada componente conectado dentro de una máscara. Se han establecido distintos umbrales para el caso de los tumores y de los edemas, ya que los edemas tienden a tener formas más irregulares y extendidas alrededor del tumor. Las condiciones establecidas son las siguientes:

1. Compacidad: Relaciona el perímetro con el área mediante (11). El círculo tiene el contorno más simple (0) mientras que 1 representa el contorno más complejo.

$$cf = 1 - \frac{4\pi A}{P^2} \quad (11)$$

Donde A es el área de la región y P es el perímetro. El componente se elimina si es mayor de 0.65 para un tumor y 0.75 para un edema.

2. Excentricidad: Excentricidad de la elipse que tiene los mismos segundos momentos que la región. La excentricidad es la relación de la distancia focal (distancia entre puntos focales) sobre la longitud del eje mayor. El valor está en el intervalo [0, 1). Cuando es 0, la elipse se convierte en un círculo. En caso de que el resultado sea mayor de 0.9 para un tumor y 0.95 para un edema, se elimina el componente de la máscara.
3. Área: En caso de que el área del componente sea mayor que el 40% del área de la máscara obtenida en 4.1.2 Eliminación del cráneo, se elimina el componente de la máscara.

b) Continuidad entre máscaras

Es importante analizar la continuidad entre máscaras para asegurarse de que las regiones segmentadas tienen sentido ya que un mismo tumor debe localizarse en la misma zona a lo largo de las imágenes de un escáner. Para ello, se calcula la coincidencia de dos máscaras contiguas a lo largo de todas las máscaras. Cuando no se encuentre ninguna coincidencia entre máscaras, se ha decidido marcar la máscara con mayor área entre todas como el tumor segmentado, eliminando las demás.

Si hay por lo menos dos máscaras conectadas, se seleccionan las máscaras desde la primera conectada hasta la última que no tenga conexión con la siguiente. En caso de que haya más de dos grupos de máscaras conectadas, se selecciona el grupo que

conecte más máscaras en su totalidad o el grupo que sume más área conectada entre todas sus máscaras.

Por último, debido a posibles errores introducidos en pasos anteriores, se concede una máscara de margen en la primera y última seleccionadas. Es decir, si la primera máscara no coincide con su anterior pero sí con la máscara que se sitúa dos posiciones antes, se considera la conexión como válida. Lo mismo se aplica para la última.

Este proceso de continuidad se expresa mejor gráficamente en la Figura 29. Se puede ver que las imágenes 3, 4 y 5 son las que más componentes consecutivos conectan y, aunque la imagen 2 no está conectada, el hecho de analizar también la que se encuentra dos posiciones antes hace que las imágenes 0 y 1 sean seleccionadas como válidas.



Figura 29: Análisis de continuidad entre máscaras.

c) Resultados de la validación

Todo este proceso descrito se puede observar en la Figura 30 donde se comparan las máscaras extraídas antes y después de la validación. Tal y como se ha descrito, el algoritmo ha segmentado alguna región en casi todas las imágenes. Aun así, tras el proceso de validación se han mantenido solamente las imágenes marcadas en verde correspondientes al tumor. Además, se ha podido mantener la continuidad entre máscaras pesar de que una ha fallado en detectar precisamente el tumor.

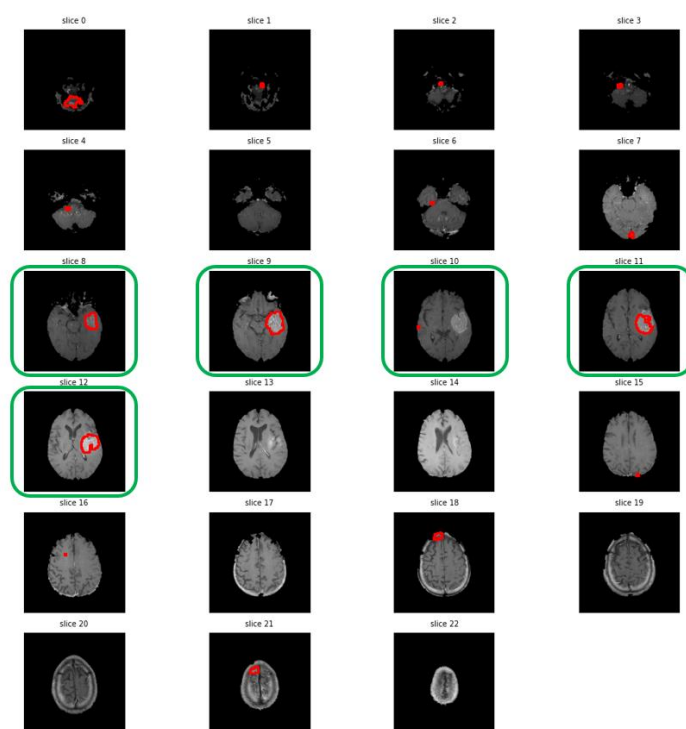


Figura 30: Resultados de la validación de las máscaras.

4.2.7. Resultados de la segmentación

Una vez se ha realizado todo el proceso de segmentación y validación, las respectivas máscaras detectadas en cada imagen tanto para el tumor como para el edema han sido extraídas como se puede ver en la Figura 31 y en la Figura 32 respectivamente. Se puede apreciar que generalmente el edema suele estar más extendido que el tumor tomando una mayor área.

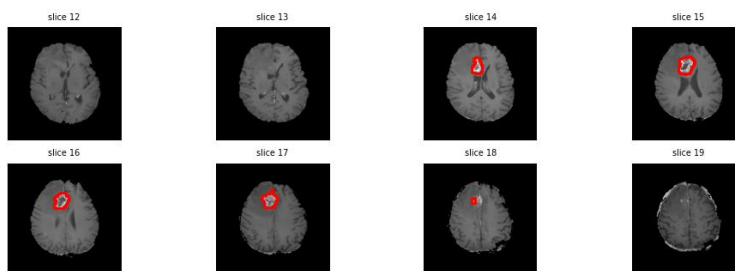


Figura 31: Resultado de la segmentación de un tumor.

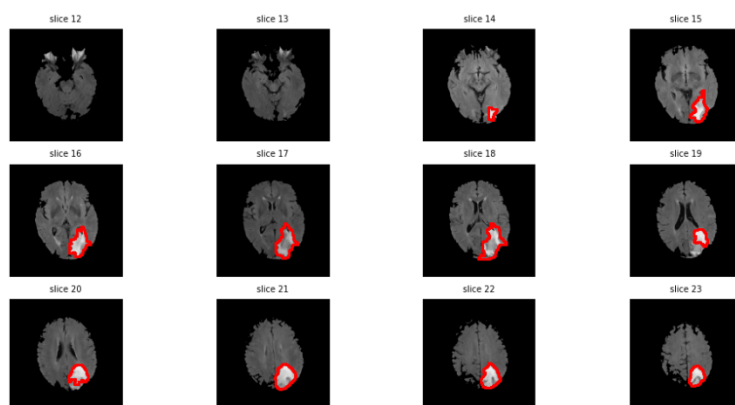


Figura 32: Resultado de la segmentación de un edema.

4.3. Extracción de características

Una vez se ha logrado segmentar tanto el tumor como el edema en distintos pacientes, hace falta extraer y calcular distintas características de las máscaras finales. Estas características deberían ser relevantes y ser indicadores del estado de la enfermedad del paciente para su posterior uso en el análisis de datos y la predicción e identificación de la situación del GBM.

Se han extraído características correspondientes a la textura, forma o posición tanto del tumor como del edema. Estas características se resumen en la Tabla 3. Cabe destacar que todas las propiedades se han calculado tanto para el tumor como para el edema. Asimismo, el volumen es la única característica que se calcula teniendo en cuenta todas las imágenes de interés. Todas las demás se extraen partiendo de la imagen que mayor volumen representa. Esto es debido a que, si se evalúan las propiedades de todas las imágenes, se obtendrían demasiadas. Además, el número total de máscaras varía según el paciente. Por lo tanto, teniendo en cuenta solo la de mayor volumen, se asegura sacar un número razonable y constante de características.

Tabla 3: Características extraídas para su posterior análisis.

CARACTERÍSTICA	NOMBRE	DESCRIPCIÓN
Volumen	Volume	Volumen total de las regiones extraídas en todas las imágenes. Se tiene en cuenta la resolución, así como el espaciado entre dos imágenes contiguas para calcular el volumen en mm ³ .
Distancias entre centroides	Dist_x Dist_y	Distancia en píxeles desde el centroide del cráneo hasta el centroide de la región de interés. Se calcula esta distancia en los ejes X e Y.
Compacidad	Compactness	Se calcula la compacidad media de todas las regiones que componen la máscara.
Valor medio de grises	Uf	El valor medio en escala de grises teniendo en cuenta las probabilidades de cada nivel presente en la imagen.
Desviación estándar	Std	Variación o dispersión de los valores.
Varianza	Var	Variabilidad de los datos respecto a su media.
Asimetría estadística	Skewness	Grado de simetría de la distribución de los datos.
Curtosis	Kur	Representa la densidad en la “cola” de una distribución.
Entropía	Ent	Cantidad de información promedio que contienen los datos.
Suavidad	Smoothness	Suavidad relativa de la intensidad en una región. Es alta para una región de intensidad constante, y baja para regiones con grandes cambios en los valores de sus niveles de intensidad.
Uniformidad	Uniformity	Consistencia en la distribución de los datos.
Matriz de coocurrencia de niveles de grises	glcm_contrast glcm_dissimilarity glcm_homogeneity glcm_energy glcm_correlation glcm_ASM	Se ha calculado la matriz de coocurrencia de niveles de grises (GLCM) tomando como referencia el píxel ubicado 5 posiciones a la derecha. Después, se han computado 6 características partiendo de los resultados: <ul style="list-style-type: none"> - Contraste: contraste de intensidad entre un píxel y la referencia en toda la imagen. - Disimilitud: diferencia entre un píxel y la referencia en toda la imagen. - Homogeneidad: cercanía espacial de la distribución de elementos a la diagonal. - Energía: tasa de cambio de intensidad de los píxeles. - Correlación: medida de correlación de un píxel respecto a su referencia en toda la imagen. - ASM: Segundo momento angular.
Momentos de Hu	Hu_{1-7}	Conjunto de 7 momentos invariable en traslación, escala y rotación calculados a partir de los momentos normales.

Teniendo en cuenta que cada parámetro se extrae para el tumor y para el edema, se obtienen un total de 50 características para cada uno de los 55 pacientes analizados con el algoritmo desarrollado. Con todo esto, se ha creado un fichero CSV donde se ha introducido una nueva fila por cada paciente.

4.3.1. Empleo de distintos tipos de escáneres

Como se han utilizado distintos tipos de escáneres de RM para la segmentación como se ha explicado en 4.2 Segmentación del *glioma* y del edema, esta información ha sido incluida en el archivo con las características extraídas. Aun así, la mejora en los resultados dada por esta combinación ha estado limitada por varios factores.

Por una parte, lo ideal es que estas imágenes estén registradas y los parámetros como el número de imágenes por escáner, la distancia entre ellas y la resolución coincidan. En ese caso, el tumor y edema se pueden superponer tras segmentarlos y realizar un análisis más robusto de todo el conjunto. Pero esto no ocurre en la base de datos seleccionada por lo que ha habido que segmentar y analizar el tumor y el edema por separado en distintos tipos de imágenes sabiendo solo que pertenecen al mismo paciente.

Por otra parte, tampoco se dispone de todos los tipos de escáneres en todos los pacientes lo que ha supuesto el uso de la vista axial como principal y la coronal o sagital en caso de que esta no esté disponible. Tanto la segmentación como la extracción de características se ha podido realizar de la misma forma ya que la presencia de las anomalías es la misma vista desde otro ángulo. No obstante, esto también dificulta la comparación de los resultados en todos los pacientes.

4.4. Análisis de los factores de riesgo

El análisis de los factores de riesgo se ha realizado, por un lado, con los metadatos proporcionados para los 189 pacientes a los que se les ha extraído muestras patológicas y, por otro lado, con la combinación de metadatos y características extraídas de las imágenes. Ambos análisis se describen a continuación.

4.4.1. Factores de riesgo basados en metadatos

Para el análisis de factores de riesgo basado únicamente en los metadatos proporcionados, se cuenta con 170 pacientes y 18 variables. Las variables se describen en la Tabla 4.

Tabla 4. Descripción de las variables disponibles para el análisis de factores de riesgo basado en los metadatos.

Nombre	Descripción	Tipo
age	Edad	Numérico
alcohol_consumption	Hábitos de consumo de alcohol	Categórico
bmi	Índice de masa corporal (IMC)	Numérico
case_id	Identificador	Categórico
ethnicity	Etnia	Categórico
race	Raza	Categórico
gender	Género	Categórico
height_in_cm	Altura	Numérico
tobacco_smoking_history	Hábitos de consumo de tabaco	Categórico

Nombre	Descripción	Tipo
number_of_pack_years_smoked	Número de años donde se haya consumido más de un paquete de tabaco al día	Numérico
tumor_site	Localización del tumor	Categorico
tumor_size_in_cm	Tamaño del tumor	Numérico
weight_in_kg	Peso	Numérico
years_high_alcohol_quantity_consumed	Número de años donde el consumo de alcohol haya superado las 2 copas al día.	Numérico
time_monitored	Tiempo de seguimiento del paciente	Numérico
cause_of_death	En caso de muerte, causa de la muerte	Categorico
vital_status	Estado del paciente (vivo/muerto)	Categorico
life_expectancy	En caso de muerte, cuantos días ha sobrevivido desde el diagnóstico.	Numérico

a) Exploratory Data Analysis (EDA)

El primer paso ha sido analizar las variables y las entradas de los pacientes, con el objetivo de eliminar aquellas que el 30% de sus entradas se corresponden a variables de tipo *Not Available* (NA), los que supone un porcentaje demasiado alto para luego realizar una imputación que se pueda ajustar a la realidad. De esta forma, se eliminan 12 pacientes y 2 variables (*ethnicity* y *race*), contando con una BD de 158 pacientes y 16 variables.

Mediante el análisis de las distribuciones de las variables se ha conseguido identificar y eliminar valores atípicos o *outliers*. También se ha procedido al análisis de la correlación entre las variables numéricas. La variable correspondiente al peso está altamente correlacionada con la altura y el IMC, pero no llega al 0.7, por lo que se decide mantener todas las variables.

Se ha analizado la relación de cada variable con las variables de estado del paciente y la supervivencia desde el diagnóstico o tiempo de monitorización, las cuales se emplean para la creación del modelo de supervivencia, con el objetivo de poder observar si hay una clara separación o tendencia. Se han empleado diagramas de caja, de barras y de dispersión, y de estas no se ha podido identificar una clara separación entre clases (Figura 33). Se han identificado ciertas tendencias, por ejemplo, se ha observado que aquellos pacientes con un peso alto han sufrido de complicaciones postoperatorias, pero esto puede no ser determinante debido al bajo porcentaje de pacientes que han sufrido de este contratiempo.

Además, se ha comprobado si el dataset está compensado para la variable del estado del paciente. El dataset contiene un 40.5% (64) de pacientes que sobreviven al seguimiento que se les ha realizado y un 59.5% (94) que no sobreviven. Aunque no esté completamente compensado, no se ha planteado el uso de técnicas para compensar la BD.

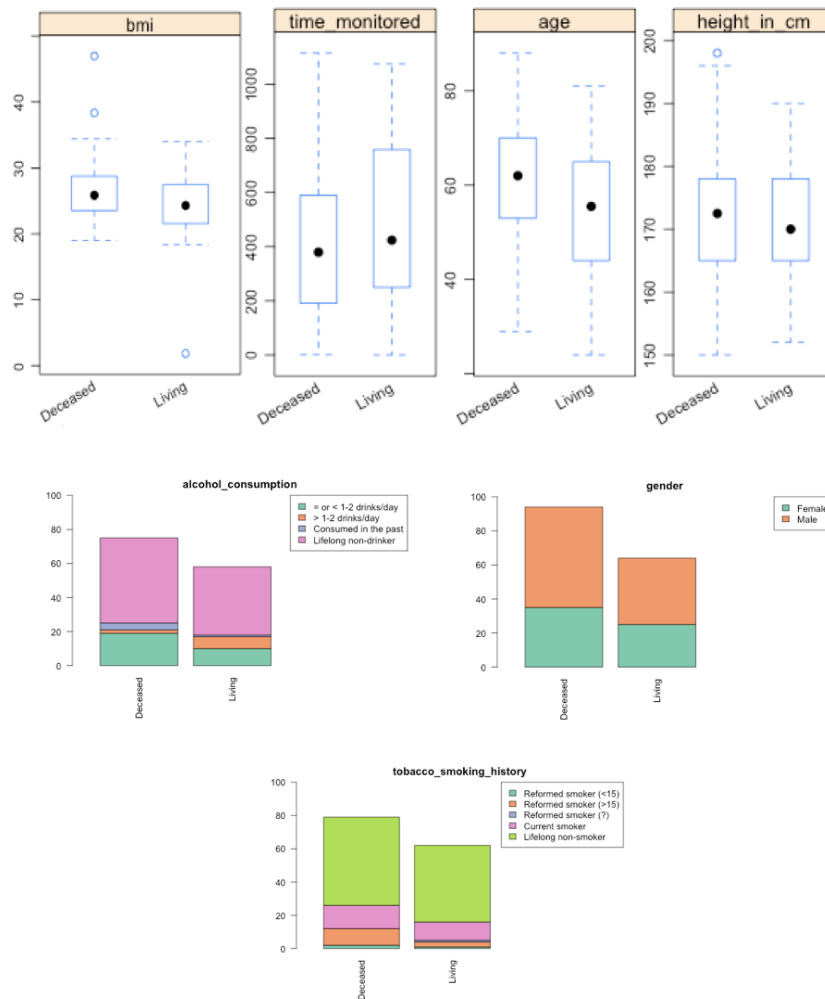


Figura 33: Gráficas extraídas para la separación de clases del análisis de factores de riesgo con los metadatos.

b) Pre-procesamiento

El alto número de valores NA que contiene la BD obliga a realizar una imputación de dichos valores. La técnica seleccionada para este proceso ha sido *Predictive Mean Matching* (PMM), debido a la capacidad de esta técnica de producir valores que se asemejan a la realidad al sustituir el valor de la variable que falta por el valor que tiene el paciente que más se le asemeja en dicha variable.

c) Modelos de supervivencia

Una vez se ha completado la BD, se procede al cálculo de las curvas de supervivencia. Se han empleado dos algoritmos diferentes para su cálculo: el algoritmo de ML RF y la regresión de riesgos proporcionales de Cox. En el caso del algoritmo de RF, la curva se determina por cada paciente del dataset y la curva de supervivencia general se extrae mediante el cálculo de la media de todas las curvas.

Se ha procedido al cálculo del modelo de supervivencia con todas las variables disponibles por cada algoritmo. De estas, se han extraído las variables más determinantes para la creación de dicha curva y se ha creado una segunda empleando

el mismo algoritmo, pero empleando únicamente la selección de variables. La Figura 34 muestra las curvas obtenidas, así como algunos de los valores de cada una de las curvas.

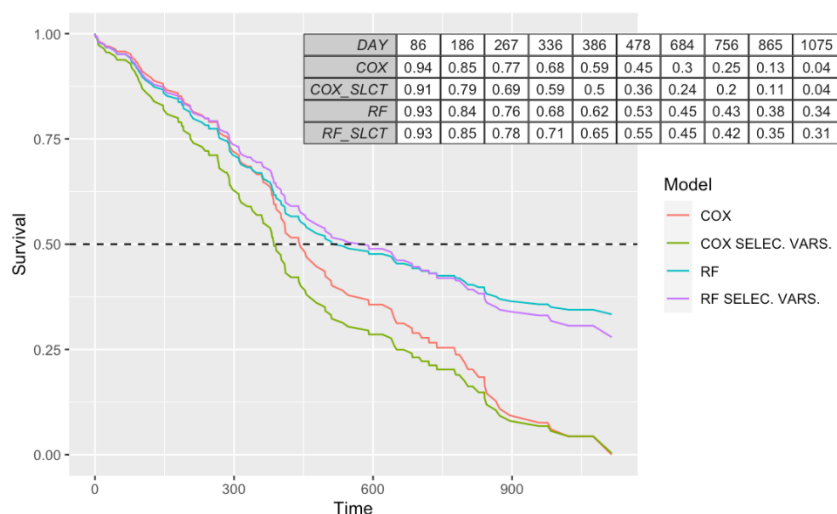


Figura 34: Curvas de supervivencia obtenidas del análisis de factores de riesgo con los metadatos.

Para determinar la relevancia de cada variable en la regresión de riesgos proporcionales de Cox se ha ejecutado una prueba estadística de significancia mediante el cálculo del *p-value*, determinando que la variable ha sido relevante en caso de que $p < 0.05$. En el algoritmo de RF se ha empleado la importancia que le asigna el propio algoritmo a cada una de las variables para determinar su relevancia. Se ha empleado la importancia asignada por permutación, variable que determina cómo mejora o empeora el modelo si se cambian los valores o se elimina la variable. Los resultados obtenidos se muestran en la Tabla 5.

Tabla 5. Resultados de relevancia por cada variable y modelo en el análisis de factores de riesgo con los metadatos.

Variable	Cox (<i>p-value</i>)		RF (importancia)	
	Todas	Selección	Todas	Selección
age	0.0035	0.0035	2.9421	10.9185
alcohol_consumption	0.0062	0.0062	3.7515	15.3904
bmi	0.0442	0.0366	-2.2684	—
gender	0.5670	—	-1.3348	—
height_in_cm	0.2233	—	-0.0040	—
tobacco_smoking_history	0.4986	—	2.6773	2.9285
number_of_pack_years_smoked	0.7269	—	-0.3636	—
tumor_site	0.1581	—	-6.2210	—
tumor_size_in_cm	0.0758	—	-1.8686	—
weight_in_kg	0.2102	—	-0.3974	—
years_high_alcohol_quantity_consumed	0.2290	—	-2.3602	—

Ambos algoritmos coinciden al señalar la edad y el consumo de alcohol, en ese orden, como los factores más determinantes para la creación de la curva de supervivencia, sin embargo, mientras que la regresión de Cox determina el IMC como

tercera variable de mayor importancia, el algoritmo RF señala el historial de consumo de tabaco como tercera variable a tener en cuenta.

También se ha realizado una prueba estadística de significancia entre los dos modelos de regresión de riesgos proporcionales de Cox, obteniendo $p = 0.1684$, lo que implica que la diferencia entre ambos modelos no es relevante. Además, implica que sólo con las variables seleccionadas se puede estimar la supervivencia del paciente de la misma manera que con todas las variables.

Como métrica para la comparación de los cuatro modelos se ha empleado la concordancia o *C-index*, obteniendo los valores mostrados en la Tabla 6 donde se observa que el mejor modelo es el creado mediante la regresión de riesgos proporcionales de Cox empleando todas las variables disponibles.

Tabla 6. Valores de la concordancia obtenidos para los modelos de supervivencia creados de los metadatos.

Cox		RF	
Todas	Selección	Todas	Selección
0.70	0.64	0.52	0.59

4.4.2. Factores de riesgo basados en características de las resonancias magnéticas y metadatos

En este caso se va a realizar el análisis a los 12 meses de seguimiento. Para ello, se cuenta con 52 pacientes para los que se han extraído las características con 53 variables. Teniendo en cuenta el ID de cada paciente, se junta con los metadatos de 170 pacientes compuestos por 18 variables. Tras esta operación, se obtienen 55 pacientes con 69 variables.

Las variables son las descritas en la Tabla 4 y las características extraídas e incluidas en el CSV como se ha explicado en 4.3 Extracción de características. Entre las características extraídas, las variables del ID del paciente (*case_id*), el tipo de RM de donde se ha obtenido el tumor (*tumor_mri_type*), y el tipo de RM de donde se ha obtenido el edema (*edema_mri_type*) son categóricas, mientras que las demás son numéricas.

a) Exploratory Data Analysis (EDA)

Debido a que algunos de los NA presentes en la BD de las características se dan en casos en los que no es posible calcular esas características porque no se ha encontrado ningún tumor o edema, por ejemplo, se sustituirán estos valores por un valor extremo como puede ser "-99999". Así, se ve que estas entradas son especiales y se podrán tener en cuenta. Esto se da cuando el volumen del edema o del tumor es 0 y tiene algún NA. Aun así, hay una excepción para los casos anteriores. Los casos en los que todos los valores son NA (incluyendo el volumen del tumor o del edema) deberán ser imputados ya que no significa que no se haya encontrado ningún tumor, sino que no se dispone de un escáner para analizarlo. También hay algún caso donde se ha detectado una

anomalía, pero existe al menos un NA. Estos NA no se cambiarán para que sean imputados después.

Tras analizar las variables y las entradas, se han eliminado las entradas con un 30% o más de variables NA y las variables con un 40% o más de variables NA, ya que es un porcentaje demasiado alto para realizar una imputación realista. Así, se han eliminado 3 pacientes y 2 variables (ethnicity y race) dejando un dataset de 49 pacientes y 67 variables.

Analizando las distribuciones de las variables no se ha encontrado ningún *outlier* que debería ser eliminado. Se ha visto que en este punto hay 1 paciente del que se han extraído características, pero no se tienen sus metadatos por lo que ha sido eliminado dejando 48 entradas. También se ha procedido al análisis de la correlación entre las variables numéricas y se van a eliminar las variables altamente correlacionadas del dataset para evitar problemas de multicolinealidad, de overfitting y errores en la creación de modelos de supervivencia. Para ello, se ha establecido un límite de correlación de 0.75 y se han eliminado 31 variables.

Ya que se quiere realizar el análisis a los 12 meses, se ha limitado el valor de la variable tiempo monitorizado (time_monitored) a 365. Igualmente, se han cambiado los valores mayores que 365 de la esperanza de vida (life_expectancy) por NA ya que un valor mayor significa que ha muerto después de los 12 meses.

Como en el caso basado en metadatos, se ha analizado la relación de cada variable con las variables de estado del paciente y la supervivencia desde el diagnóstico o tiempo de monitorización, las cuales se emplean para la creación del modelo de supervivencia, con el objetivo de poder observar si hay una clara separación o tendencia (Figura 35).

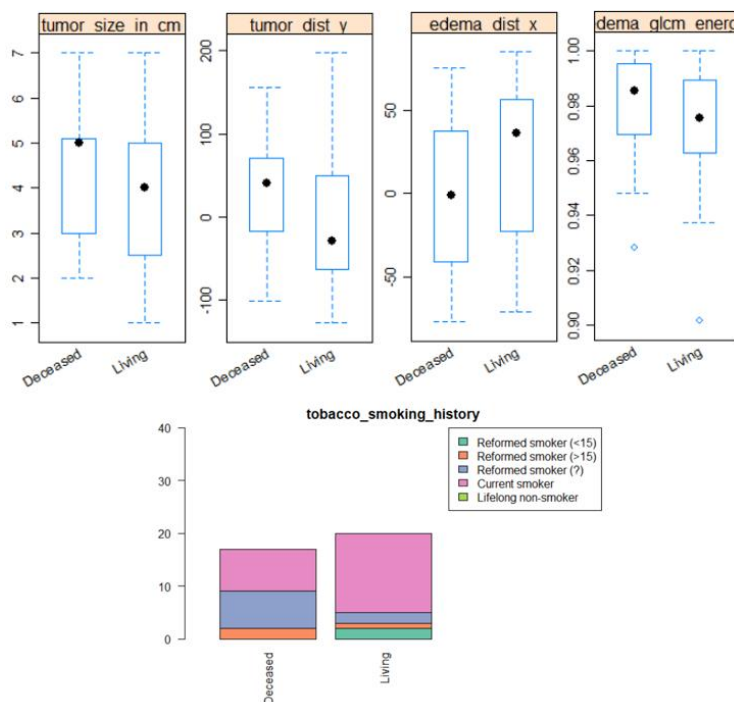


Figura 35: Gráficas extraídas para la separación de clases del análisis de factores de riesgo con los metadatos y características.

Parece que la mayoría de los que han muerto tenían un tamaño de tumor y edema algo más alto como mediana. También se han observado ciertas variaciones en la distancia en el eje Y entre centroides, el contraste de la matriz de coocurrencia del tumor y en la compacidad del tumor. También parece que los que nunca han fumado tienden a sobrevivir, pero no se observa que ninguna variable sea determinante para una predicción clara del estado del paciente. Además, se ha comprobado si el dataset está compensado para la variable del estado del paciente. El dataset contiene un 43.75% (21) de pacientes que sobreviven al seguimiento que se les ha realizado y un 56.25% (27) que no sobreviven. Aunque no esté completamente compensado, no se ha planteado el uso de técnicas para compensar la BD.

b) Pre-procesamiento

El alto número de valores NA que contiene la BD (149) obliga a realizar una imputación de dichos valores. La técnica seleccionada para este proceso ha sido la de Classification And Regression Trees (CART), basada en la clasificación y en árboles de regresión. Se ha decidido cambiar el método de imputación respecto al análisis basado en metadatos debido a que la matriz de alguna de las variables obtenida con PMM no es invertible y, por tanto, no se puede utilizar para desarrollar un modelo de regresión. Esto puede deberse a la presencia de variables linealmente dependientes o a que alguna de las variables presenta valores bastante más altos que las demás. Se ha comprobado que la normalización de las variables elimina el error por lo que se cree que la causa es la segunda razón presentada ya que, además, las variables correlacionadas ya han sido eliminadas. Teniendo esto en cuenta, se ha preferido cambiar el método de imputación a normalizar las variables con el fin de no cambiar los valores que se tienen y perder información.

Tras la imputación, el dataset no tiene ningún valor NA. Los únicos valores NA presentes ahora son los de life_expectancy, que corresponden a pacientes vivos a los 12 meses.

c) Modelos de supervivencia

Una vez se ha completado la BD, se procede al cálculo de las curvas de supervivencia empleando los algoritmos de ML RF y la regresión de riesgos proporcionales de Cox.

Se ha procedido al cálculo del modelo de supervivencia con todas las variables disponibles por cada algoritmo. De estas, se han extraído las variables más determinantes al igual que en el caso basado en metadatos para la creación de dicha curva y se ha creado una segunda empleando el mismo algoritmo, pero utilizando únicamente la selección de variables. La Figura 36 muestra las curvas obtenidas, así como algunos de los valores de cada una de las curvas.

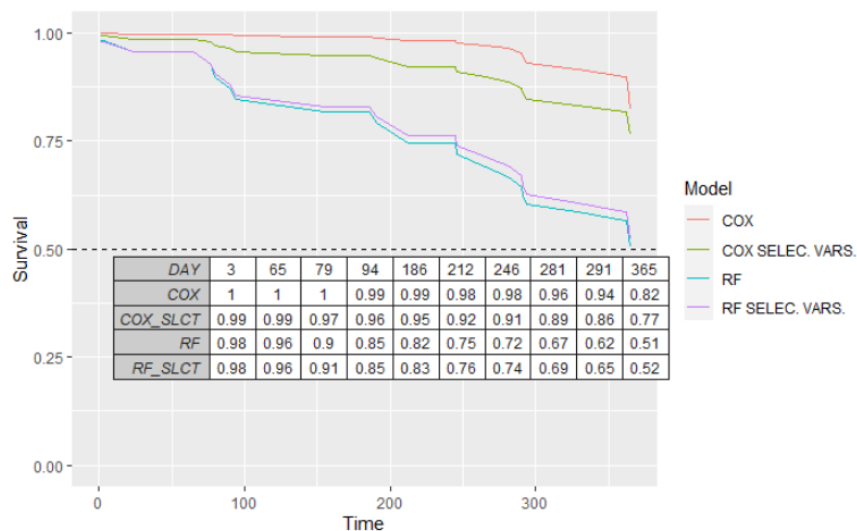


Figura 36: Curvas de supervivencia obtenidas del análisis de factores de riesgo con los metadatos y características extraídas.

Los resultados obtenidos para la relevancia de variables en cada caso se resumen en la Tabla 7. Ambos algoritmos coinciden en señalar el tercer momento de Hu del tumor y la métrica de correlación de la GLCM del edema como factores determinantes. La regresión de Cox determina además la edad como determinante. No obstante, el algoritmo RF señala las variables de historial de fumador y el tamaño del tumor en cm como unas de las más importantes junto con otras características obtenidas. Aun así, hay que tener en cuenta que se han excluido las variables del historial de fumador y el género de la regresión de Cox ya que dan un coeficiente infinito y el modelo no converge.

También se ha realizado una prueba estadística de significancia entre los dos modelos de regresión de riesgos proporcionales de Cox, obteniendo $p = 0.0675$, lo que implica que la diferencia entre ambos modelos no es relevante y se podría estimar la supervivencia del paciente sin tener que usar todas las variables asumiendo un error.

Como métrica para la comparación de los cuatro modelos se ha empleado la concordancia o *C-index*, obteniendo los valores mostrados en la Tabla 8 donde se observa que en este caso también el mejor modelo es el creado mediante la regresión de riesgos proporcionales de Cox empleando todas las variables disponibles.

Tabla 7: Resultados de relevancia por cada variable y modelo en el análisis de factores de riesgo con los metadatos y características.

Variable	Cox (<i>p-value</i>)		RF (importancia)	
	Todas	Selección	Todas	Selección
tumor_Hu_3	0.0021	0.0949	2.6285	1.6974
edema_glm_correlation	0.0095	0.2459	0.1943	-0.2751
age	0.0294	0.8106	-2.6228	—
tumor_size_in_cm	0.0582	0.3067	5.0505	6.1232
height_in_cm	0.0583	0.0063	-1.8690	—
tumor_ent	0.9124	—	1.7915	4.0602

Variable	Cox (<i>p-value</i>)		RF (importancia)	
	Todas	Selección	Todas	Selección
tumor_Hu_2	0.7458	—	1.6906	2.0785
tumor_Hu_6	0.2248	—	1.3514	0.1603
gender	—	—	1.3185	-0.8873
tumor_glm_contrast	0.7396	—	1.1920	-1.4654
tobacco_smoking_history	—	—	0.6475	3.6388
tumor_smoothness	0.2800	—	0.4423	3.1936

Tabla 8: Valores de la concordancia obtenidos para los modelos de supervivencia creados de los metadatos y características.

Cox		RF	
Todas	Selección	Todas	Selección
0.82	0.69	0.36	0.51

4.5. Modelo de predicción de supervivencia

Para la predicción de la supervivencia de un paciente a 12 meses del diagnóstico se ha empleado la BD resultante de la creación de la curva de supervivencia con los metadatos y las características extraídas de las RM descrita en el apartado 4.4.2 Factores de riesgo basados en características de las resonancias magnéticas y metadatos. De esta forma, y partiendo de esta BD, se han dividido los 48 pacientes en dos dataset distintos, uno para el entrenamiento del modelo, el cual contiene el 80% de las entradas y otro para la validación, que contiene el 20% de los pacientes.

Partiendo de esto, se han creado 3 vistas distintas de la BD. La primera vista se corresponde a la obtenida en el apartado 4.4.2 Factores de riesgo basados en características de las resonancias magnéticas y metadatos, donde se han eliminado las variables numéricas con alta correlación. Para la segunda vista, se ha aplicado el método PCA a las variables numéricas para la reducción de la dimensionalidad del dataset al haber una alta proporción de variables con respecto a pacientes. En esta vista se consigue pasar de 35 variables a 17 manteniendo el 95% de la variabilidad de la BD. Para la tercera y última vista se han seleccionado las variables determinadas como factores de riesgo en el apartado 4.4 Análisis de los factores de riesgo y se ha creado una BD con ellas.

Con cada una de las vistas se han creado 5 modelos distintos mediante 5 algoritmos de ML. Los algoritmos empleados han sido KNN, SVM, RF, *Boosting* (combinación de distintos algoritmos sencillos) y Regresión Logística. Estas métricas han sido evaluadas con la opción de la curva ROC.

Cada algoritmo ha sido entrenado empleando la opción *repeated cross-validation*, en la que la *cross-validation* se ha ejecutado 5 *folds* y 3 veces para cada modelo con el objetivo de obtener más datos para la comparación de modelos.

4.5.1. Comparación de los modelos

De todas las métricas disponibles, se han seleccionado Accuracy y LogLoss para evaluar la capacidad de los modelos de predecir y ver cómo se penaliza en caso de que no predigan correctamente con la métrica LogLoss. Los valores obtenidos para ambas métricas se muestran en la Figura 37.

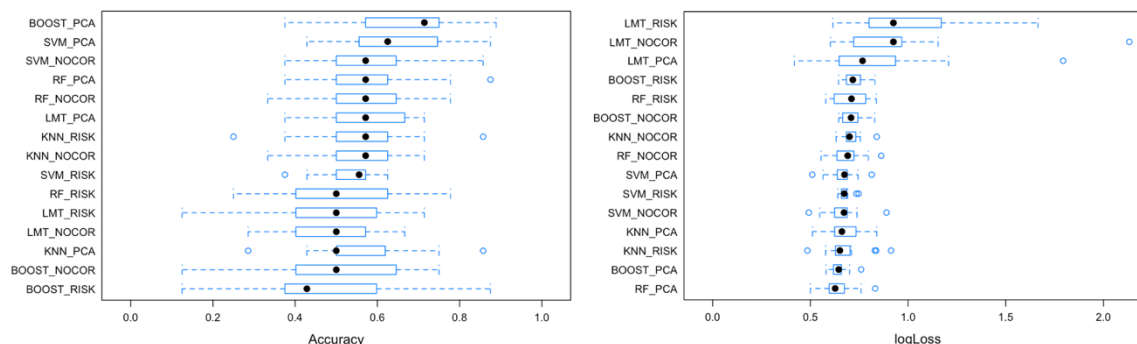


Figura 37: Valores de las métricas Accuracy y LogLoss para los modelos creados.

Empleando únicamente las gráficas se observa que el algoritmo de *Boosting* aplicado al dataset al cual se le ha aplicado PCA es el que mejor *Accuracy* y *LogLoss* tiene, pero para obtener una comparación más representativa se han realizado pruebas de estadística de significancia para ambas métricas. De estas pruebas se han extraído diferentes conclusiones sobre los modelos y su rendimiento.

- De todos los modelos creados con el dataset en dónde PCA ha sido aplicado, *Boosting* ha sido el mejor en lo que respecta a *Accuracy* y a *LogLoss*. El algoritmo de Regresión Logística es el que peor rendimiento tiene en lo que corresponde a *LogLoss*, siendo la diferencia con el resto significativa.
- De los modelos creados con las variables no correlacionadas el mejor en lo que respecta al *Accuracy* han sido RF y SVM, pero los valores del *p-value* muestran que la diferencia entre ellos es significativa, siendo SVM el mejor de los dos. SVM es también el algoritmo que mejor funciona respecto a la métrica *LogLoss*.
- En lo que respecta a los algoritmos en los cuales la base de datos empleada ha sido generada únicamente con los factores de riesgo, KNN y Regresión Logística son los que mejor rendimiento presentan para la métrica de *Accuracy*, sin ningún cambio representativo entre ambos modelos, aunque Regresión Logística es el algoritmo que mayor penalización recibe en la métrica *LogLoss*.

En general, el valor del *Accuracy* es bastante bajo para todos los casos, siendo el mejor el algoritmo *Boosting* para la BD al cual se le ha aplicado PCA. A su vez, este modelo es el que menos es penalizado mediante la métrica *LogLoss*. El segundo algoritmo mejor posicionado es el SVM para la BD de PCA. De esta forma, se propone el uso de estos dos modelos para la validación. Dado que PCA puede dificultar la interpretación de los resultados, se ha decidido emplear también el modelo entrenado mediante SVM para la BD de las variables no correlacionadas.

5. EVALUACIÓN

Se presenta la evaluación de las distintas implementaciones realizadas.

5.1. Algoritmo para la eliminación del cráneo

Se ha creado un pequeño programa que compara la máscara resultante de la implementación del algoritmo para la eliminación del cráneo con una máscara que se considera como la realidad (*ground truth*).

5.1.1. Dataset

Para poder evaluar el algoritmo correctamente, se necesita de la RM original y la máscara que se considerará como la realidad, es decir, cuál sería la máscara resultante de una extracción perfecta. Dado que el dataset seleccionado para el desarrollo del proyecto no cuenta con la máscara de la extracción perfecta, se ha empleado un dataset [92] disponible en la Internet Brain Segmentation Repository (IBSR) de la página Neuroimaging Tools & Resources Collaboratory. Cuenta con imágenes de RM de 18 pacientes distintos, y todos cuentan con el volumen original, la máscara resultante de la extracción y el volumen con el cráneo eliminado.

5.1.2. Resultados

Se han calculado dos tipos de métricas. Por un lado, se calcula el error de la segmentación obteniendo el número de falsos positivos (FPR) y de falsos negativos (FNR). Estas métricas se obtienen dividiendo el número de vóxeles clasificados erróneamente como cerebro o no, respectivamente, entre el número de vóxeles que pertenecen al cerebro en el volumen. Además, también se obtiene el número de verdaderos positivos (TPR), dividiendo el número de vóxeles clasificados como cerebro correctamente entre el número total de vóxeles asociados al cerebro en el *ground truth*.

Por otro lado, se calcula el índice de *Jaccard* para obtener el grado de similitud entre los dos volúmenes. Este índice se calcula mediante la división de la intersección y la unión de ambos volúmenes. La media y las desviaciones de las métricas se muestra en la Tabla 9.

Tabla 9. Media de las métricas obtenidas para la evaluación del algoritmo de extracción del cerebro.

TPR	FPR	FNR	<i>Jaccard</i>
0,869 ± 0,161	0,320 ± 0,153	0,131 ± 0,161	0,662 ± 0,135

Los resultados obtenidos son buenos exceptuando en los sujetos con id IBSR_08, IBSR_10 e IBSR_11 (los resultados completos se encuentran en el APÉNDICE A). Esto se debe a que en los volúmenes originales se observa una peor calidad de la imagen, y ese ruido no permite el correcto rendimiento del algoritmo. Además, el FPR se considera que es demasiado alto para una correcta extracción del tumor.

5.2. Modelo final de Machine Learning

Para crear los modelos finales y evaluar su rendimiento con la BD de evaluación, se crean los modelos utilizando todos los datos disponibles en la BD de entrenamiento. Esto se traduce en no utilizar el método *cross-validation* al crear los modelos. Además, se utilizan los mejores parámetros obtenidos al entrenar los modelos anteriores para asegurar que el modelo creado es el mejor de entre todas las opciones dadas en la ejecución anterior. Esto se añade como parámetro *tuneGrid* de la función de entrenamiento del modelo.

Como se concluye en el apartado 4.5.1 Comparación de los modelos, se crean 3 modelos distintos. Dado que dos de los modelos emplean datos a los cuales PCA ha sido aplicado, la BD de validación deberá sufrir la misma transformación.

La Figura 38 se muestran las matrices de confusión obtenidas para los 3 modelos, siendo a) el algoritmo de *Boosting* con la BD de PCA, b) el algoritmo SVM para la BD de PCA y c) el algoritmo SVM para la BD sin variables correlacionadas.

Observando los resultados disponibles, se observa que el mejor caso es el algoritmo de *Boosting* para la BD de PCA, obteniendo un Accuracy de 0.67 con un TPR del 0.75 y un TNR del 0.6. En los otros algoritmos, hay una prevalencia de la predicción del nivel *Living* para casi todos los pacientes.

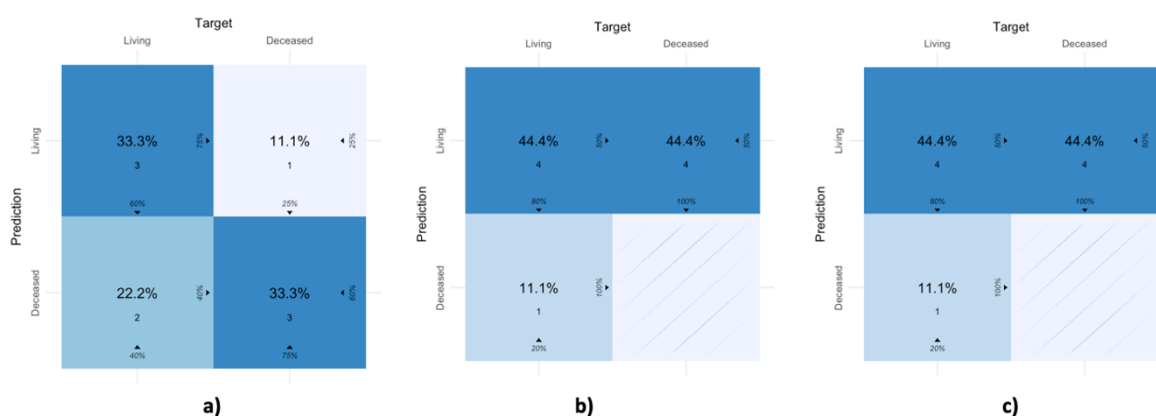


Figura 38: Matrices de confusión obtenidas en la validación de los modelos.

6. CONCLUSIONES

Tras finalizar el proyecto y haber analizado el trabajo realizado, se han sacado varias conclusiones.

Antes que nada, se ha logrado encontrar una BD válida con información de RM de suficiente volumen, 66 pacientes en concreto. De estos pacientes se han podido emplear un total de 53 para el análisis empleando las tres vistas disponibles (axial, sagital y coronal).

Se ha realizado un análisis extenso del estado del arte para evaluar la situación actual y enmarcar el proyecto teniendo en cuenta los recursos y el tiempo disponibles. Por ejemplo, se ha observado que nuevas técnicas de DL son las que actualmente ofrecen mejores resultados para la segmentación, pero han sido descartadas por su mayor complejidad.

Se han cumplido los objetivos principales definidos, los cuales determinaban la necesidad de conseguir la extracción del tumor de las RM, el análisis de factores de riesgo que marcan la supervivencia del paciente a 3 años y la creación de un modelo que permita la predicción de la supervivencia del paciente a un año.

La correcta extracción del tumor ha sido un paso determinante dónde los errores afectan enormemente a las características extraídas de estas, y de esta forma, a los modelos de supervivencia y predicción creados empleando estos datos. Sin embargo, no ha sido posible comprobar si el sistema cumple con el requisito de realizar la segmentación con una precisión de $\pm 3-4$ mm ya que no se ha hecho una evaluación de este algoritmo.

De la misma forma, el pre-procesamiento de las RM ha sido un paso clave, donde se ha logrado una correcta normalización de las imágenes RM y extracción del cerebro de ellas. La evaluación de este algoritmo ha mostrado unos resultados que, aun siendo relativamente buenos, habría que mejorar ya que el proceso de segmentación del GBM requiere de una previa segmentación del cerebro lo más exacta posible dado que el algoritmo puede llegar a confundir el cráneo con parte del tumor. Ha sido posible desarrollar este algoritmo para que sea posible de aplicar para RM de tipo T1 con contraste, FLAIR y T2.

Con el objetivo de mejorar las predicciones y para completar la extracción del tumor, se ha incluido la extracción del edema a partir de RM de tipo T2 y FLAIR para añadir las características extraídas de estos a los modelos.

El análisis de factores de riesgo ha mostrado la relevancia de hábitos como el consumo de tabaco y alcohol y características como el IMC y la edad del paciente. Además, juntando estas características con las obtenidas de las RM, muestra que los momentos de Hu, distintas métricas obtenidas de la GLCM como la correlación y el contraste, u otras características como la entropía y la suavidad de la región segmentada también son indicadores relevantes de la situación de la enfermedad.

Respecto al requisito técnico de obtener un 0.7 en el *Accuracy*, TPR y TNR, no se ha logrado llegar a ese límite establecido, aunque los resultados quedan cerca de este umbral. Además, el modelo de predicción se ha creado aplicando PCA en la BD por lo que los resultados pierden su interpretabilidad.

El código desarrollado ha sido comprobado mediante herramientas de análisis estático, mostrando la presencia de 0 *bugs* pero varios *code smells*. Además, se han seguido las buenas prácticas establecidas por PEP8. Los resultados se muestran en el APÉNDICE B RESULTADOS DEL ANÁLISIS ESTÁTICO

7. LÍNEAS FUTURAS

En este apartado se presentan las mejoras y siguientes pasos a desarrollar para la continuación del proyecto.

A pesar de haber obtenido una BD válida para el propósito del proyecto, un mayor número de pacientes hubiese aumentado la fiabilidad de los resultados posibilitando sacar conclusiones más completas y genéricas.

Una de las tareas a realizar es la mejora del algoritmo de eliminación del cráneo, cuya fiabilidad queda en duda al observar los resultados obtenidos en la evaluación de éste. De la misma forma, al no haber evaluado la segmentación del tumor y el edema, su fiabilidad también queda en duda, por lo que será necesaria la evaluación de los algoritmos y la mejora de estos en caso de no obtener buenos resultados.

Aunque se haya hecho uso de distintos tipos de RM, no se ha podido realizar un registro entre ellos para establecer así una relación entre tumor y edema. Se considera que un registro entre estas imágenes mejoraría la eficacia del algoritmo.

Otra tarea que realizar en relación con el código desarrollado es la corrección de los *code smells* obtenidos en análisis estático llevado a cabo, los cuales no han sido modificados.

El código desarrollado es una demo el cual contiene las funcionalidades necesarias para llevar a cabo todo el proceso definido en el proyecto, pero no sería aplicable, tal y como está, para su uso. Debido a esto, una tarea futura implicaría el desarrollar una aplicación para el uso de profesionales médicos que cuente con una interfaz gráfica mediante la cual permitir la visualización de los datos de los pacientes, las RM capturadas, las segmentaciones del tumor y el edema extraídos por los algoritmos, y las predicciones de los modelos de ML. Asimismo, esta aplicación podría estar conectada a un sistema de computación en la nube dónde se procesen y guarden los datos y la aplicación sea una herramienta de acceso a los datos guardados en él.

Finalmente, se propone el uso de técnicas de DL para mejorar el rendimiento los distintos apartados del proceso de detección y predicción del GBM. Los buenos resultados observados en la literatura sobre la aplicación de técnicas de DL para la extracción del tumor y la predicción de la supervivencia del paciente (descritas en 2.3.2 Nuevas técnicas de Deep Learning) invitan a la aplicación de dichas técnicas para la mejora de la fiabilidad del sistema.

8. BIBLIOGRAFÍA

- [1] «¿Qué es el cáncer? - Instituto Nacional del Cáncer», feb. 09, 2015. <https://www.cancer.gov/espanol/cancer/naturaleza/que-es> (accedido nov. 13, 2020).
- [2] F. Hanif, K. Muzaffar, kahkashan Perveen, S. Malhi, y S. Simjee, «Glioblastoma Multiforme: A Review of its Epidemiology and Pathogenesis through Clinical Presentation and Treatment», *Asian Pac. J. Cancer Prev.*, vol. 18, n.º 1, ene. 2017, doi: 10.22034/APJCP.2017.18.1.3.
- [3] D. Armocida, A. Pesce, F. Di Giammarco, A. Frati, A. Santoro, y M. Salvati, «Long Term Survival in Patients Suffering from Glioblastoma Multiforme: A Single-Center Observational Cohort Study», *Diagnostics*, vol. 9, n.º 4, p. 209, nov. 2019, doi: 10.3390/diagnostics9040209.
- [4] «Early Detection Can Be Key to Surviving a Brain Tumor», *Weill Cornell Brain and Spine Center*, ago. 08, 2016. <https://weillcornellbrainandspine.org/early-detection-can-be-key-surviving-brain-tumor> (accedido dic. 09, 2020).
- [5] P. J. Kelly, «Gliomas: Survival, origin and early detection», *Surg. Neurol. Int.*, vol. 1, dic. 2010, doi: 10.4103/2152-7806.74243.
- [6] «Glioblastoma (GBM) – American Brain Tumor Association». https://www.abta.org/tumor_types/glioblastoma-gbm/ (accedido dic. 09, 2020).
- [7] «Early Detection», *Glioblastoma Foundation*. <https://glioblastomafoundation.org/research/early-detection> (accedido nov. 13, 2020).
- [8] Committee on Diagnostic Error in Health Care, Board on Health Care Services, Institute of Medicine, y The National Academies of Sciences, Engineering, and Medicine, *Improving Diagnosis in Health Care*. Washington, D.C.: National Academies Press, 2015, p. 21794. doi: 10.17226/21794.
- [9] H. Singh, A. Meyer, y E. Thomas, «The frequency of diagnostic errors in outpatient care: Estimations from three large observational studies involving US adult populations», *BMJ Qual. Saf.*, vol. 23, abr. 2014, doi: 10.1136/bmjqs-2013-002627.
- [10] «Assessing Concordance With Watson for Oncology, a Cognitive Computing Decision Support System for Colon Cancer Treatment in Korea | JCO Clinical Cancer Informatics». <https://ascopubs.org/doi/full/10.1200/CCI.17.00109> (accedido dic. 10, 2020).
- [11] W. Taal, J. Bromberg, y M. Bent, «Chemotherapy in glioma», *CNS Oncol.*, vol. 4, pp. 1-14, abr. 2015, doi: 10.2217/cns.15.2.
- [12] T. C. Booth, M. Williams, A. Luis, J. Cardoso, K. Ashkan, y H. Shuaib, «Machine learning and glioma imaging biomarkers», *Clin. Radiol.*, vol. 75, n.º 1, pp. 20-32, ene. 2020, doi: 10.1016/j.crad.2019.07.001.
- [13] C. Okuma y R. Fernández, «EVALUACIÓN DE GLIOMAS POR TÉCNICAS AVANZADAS DE RESONANCIA MAGNÉTICA», *Rev. Médica Clínica Las Condes*, vol. 28, n.º 3, pp. 360-377, may 2017, doi: 10.1016/j.rmcl.2017.05.005.
- [14] A. A. Abd-Elghany *et al.*, «Radiological characteristics of glioblastoma multiforme using CT and MRI examination», *J. Radiat. Res. Appl. Sci.*, vol. 12, n.º 1, pp. 289-293, ene. 2019, doi: 10.1080/16878507.2019.1655864.

- [15] C. Dupont, N. Betrouni, N. Reyns, y M. Vermandel, «On Image Segmentation Methods Applied to Glioblastoma: State of Art and New Trends», *IRBM*, vol. 37, n.º 3, pp. 131-143, jun. 2016, doi: 10.1016/j.irbm.2015.12.004.
- [16] Y. Wan, R. Rahmat, y S. J. Price, «Deep learning for glioblastoma segmentation using preoperative magnetic resonance imaging identifies volumetric features associated with survival», *Acta Neurochir. (Wien)*, vol. 162, n.º 12, pp. 3067-3080, dic. 2020, doi: 10.1007/s00701-020-04483-7.
- [17] S. Bauer, R. Wiest, L.-P. Nolte, y M. Reyes, «A survey of MRI-based medical image analysis for brain tumor studies», *Phys. Med. Biol.*, vol. 58, n.º 13, pp. R97-R129, jun. 2013, doi: 10.1088/0031-9155/58/13/R97.
- [18] K. Sheela.V y D. Babu, «Pre-Processing Technique for Brain Tumor Detection and Segmentation», 2015. /paper/Pre-Processing-Technique-for-Brain-Tumor-Detection-Sheela.V.-Babu/75dbea602733c5120552162fb948371b8127a940 (accedido ene. 20, 2021).
- [19] S. Suhas y C. R. Venugopal, «MRI image preprocessing and noise removal technique using linear and nonlinear filters», en *2017 International Conference on Electrical, Electronics, Communication, Computer, and Optimization Techniques (ICEECOT)*, dic. 2017, pp. 1-4. doi: 10.1109/ICEECOT.2017.8284595.
- [20] M. Sunaryo y M. Hariadi, «Preprocessing on Digital Image using Histogram Equalization: An Experiment Study on MRI Brain Image», 2016. /paper/Preprocessing-on-Digital-Image-using-Histogram-%3A-An-Sunaryo-Hariadi/3fce14df2185e2f1ffa93f7610e8a7ec8e536ae6 (accedido ene. 21, 2021).
- [21] A. Makropoulos *et al.*, «Automatic Whole Brain MRI Segmentation of the Developing Neonatal Brain», *IEEE Trans. Med. Imaging*, vol. 33, n.º 9, pp. 1818-1831, sep. 2014, doi: 10.1109/TMI.2014.2322280.
- [22] F. Zhou, Y. Zhuang, H. Gong, J. Zhan, M. Grossman, y Z. Wang, «Resting State Brain Entropy Alterations in Relapsing Remitting Multiple Sclerosis», *PLOS ONE*, vol. 11, n.º 1, p. e0146080, ene. 2016, doi: 10.1371/journal.pone.0146080.
- [23] P. Tanskanen *et al.*, «Hippocampus and amygdala volumes in schizophrenia and other psychoses in the Northern Finland 1966 birth cohort», *Schizophr. Res.*, vol. 75, n.º 2-3, pp. 283-294, jun. 2005, doi: 10.1016/j.schres.2004.09.022.
- [24] J. Leote, R. G. Nunes, L. Cerqueira, R. Loução, y H. A. Ferreira, «Reconstruction of white matter fibre tracts using diffusion kurtosis tensor imaging at 1.5T: Pre-surgical planning in patients with gliomas», *Eur. J. Radiol. Open*, vol. 5, pp. 20-23, 2018, doi: 10.1016/j.ejro.2018.01.002.
- [25] S. Roy y P. Maji, «A simple skull stripping algorithm for brain MRI», en *2015 Eighth International Conference on Advances in Pattern Recognition (ICAPR)*, Kolkata, India, ene. 2015, pp. 1-6. doi: 10.1109/ICAPR.2015.7050671.
- [26] Y. Wu, Z. Zhao, W. Wu, Y. Lin, y M. Wang, «Automatic glioma segmentation based on adaptive superpixel», *BMC Med. Imaging*, vol. 19, n.º 1, p. 73, ago. 2019, doi: 10.1186/s12880-019-0369-6.
- [27] E. Lotan, R. Jain, N. Razavian, G. M. Fatterpekar, y Y. W. Lui, «State of the Art: Machine Learning Applications in Glioma Imaging», *Am. J. Roentgenol.*, vol. 212, n.º 1, pp. 26-37, oct. 2018, doi: 10.2214/AJR.18.20218.
- [28] I. Despotović, B. Goossens, y W. Philips, «MRI Segmentation of the Human Brain: Challenges, Methods, and Applications», *Comput. Math. Methods Med.*, vol. 2015, pp. 1-23, 2015, doi: 10.1155/2015/450341.

- [29] M. Ben Abdallah, M. Blonski, S. Wantz-Mezieres, Y. Gaudeau, L. Taillandier, y J.-M. Moureaux, «Statistical evaluation of manual segmentation of a diffuse low-grade glioma MRI dataset», *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Annu. Int. Conf.*, vol. 2016, pp. 4403-4406, ago. 2016, doi: 10.1109/EMBC.2016.7591703.
- [30] A. Bianchi, J. V. Miller, E. T. Tan, y A. Montillo, «BRAIN TUMOR SEGMENTATION WITH SYMMETRIC TEXTURE AND SYMMETRIC INTENSITY-BASED DECISION FORESTS», *Proc. IEEE Int. Symp. Biomed. Imaging Nano Macro IEEE Int. Symp. Biomed. Imaging*, vol. 2013, pp. 748-751, abr. 2013, doi: 10.1109/ISBI.2013.6556583.
- [31] M. B. Cuadra, C. Pollo, A. Bardera, O. Cuisenaire, J.-G. Villemure, y J.-P. Thiran, «Atlas-based segmentation of pathological MR brain images using a model of lesion growth», *IEEE Trans. Med. Imaging*, vol. 23, n.º 10, pp. 1301-1314, oct. 2004, doi: 10.1109/TMI.2004.834618.
- [32] A. Essadiki, E. Ouabida, y A. Bouzid, «Brain tumor segmentation with Vander Lugt correlator based active contour», *Comput. Methods Programs Biomed.*, vol. 160, pp. 103-117, jul. 2018, doi: 10.1016/j.cmpb.2018.04.004.
- [33] J. Long *et al.*, «Cascaded hybrid residual U-Net for glioma segmentation», *Multimed. Tools Appl.*, vol. 79, n.º 33, pp. 24929-24947, sep. 2020, doi: 10.1007/s11042-020-09210-z.
- [34] «Survival Analysis». https://sphweb.bumc.bu.edu/otlt/mph-modules/bs/bs704_survival/BS704_Survival_print.html (accedido may 24, 2021).
- [35] L. L. Johnson y J. H. Shih, «CHAPTER 20 - An Introduction to Survival Analysis», en *Principles and Practice of Clinical Research (Second Edition)*, J. I. Gallin y F. P. Ognibene, Eds. Burlington: Academic Press, 2007, pp. 273-282. doi: 10.1016/B978-012369440-9/50024-4.
- [36] F. E. Harrell Jr, R. M. Califf, D. B. Pryor, K. L. Lee, y R. A. Rosati, «Evaluating the Yield of Medical Tests», *JAMA*, vol. 247, n.º 18, pp. 2543-2546, may 1982, doi: 10.1001/jama.1982.03320430047030.
- [37] Stephanie, «C-Statistic: Definition, Examples, Weighting and Significance», *Statistics How To*, ago. 28, 2016. <https://www.statisticshowto.com/c-statistic/> (accedido may 26, 2021).
- [38] M. Schmid, M. N. Wright, y A. Ziegler, «On the use of Harrell's C for clinical risk prediction via random survival forests», *Expert Syst. Appl.*, vol. 63, pp. 450-459, nov. 2016, doi: 10.1016/j.eswa.2016.07.018.
- [39] «Ascent of machine learning in medicine», *Nat. Mater.*, vol. 18, n.º 5, Art. n.º 5, may 2019, doi: 10.1038/s41563-019-0360-1.
- [40] K. Kourou, T. P. Exarchos, K. P. Exarchos, M. V. Karamouzis, y D. I. Fotiadis, «Machine learning applications in cancer prognosis and prediction», *Comput. Struct. Biotechnol. J.*, vol. 13, pp. 8-17, ene. 2015, doi: 10.1016/j.csbj.2014.11.005.
- [41] J. A. M. Sidey-Gibbons y C. J. Sidey-Gibbons, «Machine learning in medicine: a practical introduction», *BMC Med. Res. Methodol.*, vol. 19, n.º 1, p. 64, dic. 2019, doi: 10.1186/s12874-019-0681-4.
- [42] S. Anto, «Supervised Machine Learning Approaches for Medical Data Set Classification-A Review», ene. 2011.

- [43] «Random Forests Classifiers in Python», *DataCamp Community*, may 16, 2018. <https://www.datacamp.com/community/tutorials/random-forests-classifier-python> (accedido ene. 27, 2021).
- [44] R. Bhatia, «Top 6 Regression Algorithms Used In Analytics & Data Mining», *Analytics India Magazine*, sep. 18, 2017. <https://analyticsindiamag.com/top-6-regression-algorithms-used-data-mining-applications-industry/> (accedido ene. 27, 2021).
- [45] A. Ye, «5 Regression Algorithms You Need to Know — Theory & Implementation», *Medium*, nov. 18, 2020. <https://medium.com/analytics-vidhya/5-regression-algorithms-you-need-to-know-theory-implementation-37993382122d> (accedido ene. 27, 2021).
- [46] S. Damilola, «A Review of Unsupervised Artificial Neural Networks with Applications», *Int. J. Comput. Appl.*, vol. 181, n.º 40, pp. 22-26, feb. 2019, doi: 10.5120/ijca2019918425.
- [47] K. G. Kim, «Book Review: Deep Learning», *Healthc. Inform. Res.*, vol. 22, n.º 4, p. 351, 2016, doi: 10.4258/hir.2016.22.4.351.
- [48] S. Albawi, T. A. Mohammed, y S. Al-Zawi, «Understanding of a convolutional neural network», en *2017 International Conference on Engineering and Technology (ICET)*, ago. 2017, pp. 1-6. doi: 10.1109/ICEngTechnol.2017.8308186.
- [49] K. O'Shea y R. Nash, «An Introduction to Convolutional Neural Networks», *ArXiv151108458 Cs*, dic. 2015, Accedido: ene. 22, 2021. [En línea]. Disponible en: <http://arxiv.org/abs/1511.08458>
- [50] S. Mutuvi, «Introduction to Machine Learning Model Evaluation», *Medium*, feb. 05, 2020. <https://heartbeat.fritz.ai/introduction-to-machine-learning-model-evaluation-fa859e1b2d7f> (accedido dic. 14, 2020).
- [51] D. Johari, T. K. A. Rahman, y P. Dr. I. Musirin, «ANN Model Selection and Performance Evaluation for Lightning Prediction System», jul. 2009.
- [52] K. Nighania, «Various ways to evaluate a machine learning models performance», *Medium*, ene. 30, 2019. <https://towardsdatascience.com/various-ways-to-evaluate-a-machine-learning-models-performance-230449055f15> (accedido dic. 14, 2020).
- [53] «Imbalanced Classification | Handling Imbalanced Data using Python», *Analytics Vidhya*, jul. 23, 2020. <https://www.analyticsvidhya.com/blog/2020/07/10-techniques-to-deal-with-class-imbalance-in-machine-learning/> (accedido ene. 27, 2021).
- [54] B. Rocca, «Handling imbalanced datasets in machine learning», *Medium*, mar. 30, 2019. <https://towardsdatascience.com/handling-imbalanced-datasets-in-machine-learning-7a0e84220f28> (accedido ene. 27, 2021).
- [55] D. Faggella, «Machine Learning for Medical Diagnostics - 4 Current Applications», *Emerj*. <https://emerj.com/ai-sector-overviews/machine-learning-medical-diagnostics-4-current-applications/> (accedido dic. 10, 2020).
- [56] «Babylon Health UK - The Online Doctor and Prescription Services App», *Babylon Health*. <https://www.babylonhealth.com/> (accedido dic. 11, 2020).
- [57] «Better Together». <https://hms.harvard.edu/news/better-together> (accedido dic. 11, 2020).

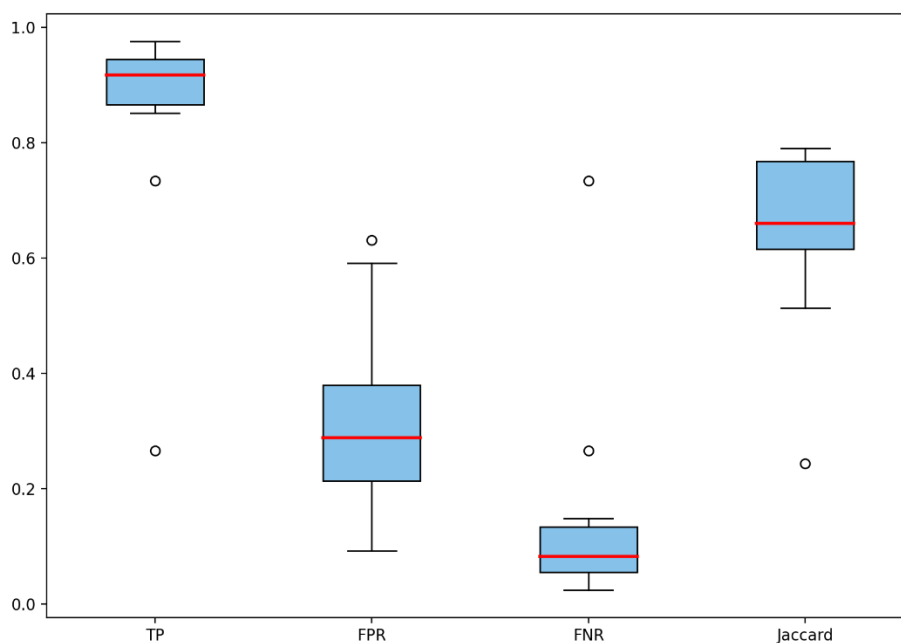
- [58] «Face2Gene Technology - How It Works», *Face2Gene*. <https://www.face2gene.com/technology-facial-recognition-feature-detection-phenotype-analysis/> (accedido dic. 11, 2020).
- [59] A. Esteva *et al.*, «Dermatologist-level classification of skin cancer with deep neural networks», *Nature*, vol. 542, n.º 7639, Art. n.º 7639, feb. 2017, doi: 10.1038/nature21056.
- [60] J. Valdebenito y F. Medina, «Machine learning approaches to study glioblastoma: A review of the last decade of applications», *Cancer Rep.*, vol. 2, n.º 6, p. e1226, 2019, doi: <https://doi.org/10.1002/cnr2.1226>.
- [61] H. Sotoudeh *et al.*, «Artificial Intelligence in the Management of Glioma: Era of Personalized Medicine», *Front. Oncol.*, vol. 9, 2019, doi: 10.3389/fonc.2019.00768.
- [62] Z. Qian *et al.*, «Differentiation of glioblastoma from solitary brain metastases using radiomic machine-learning classifiers», *Cancer Lett.*, vol. 451, pp. 128-135, jun. 2019, doi: 10.1016/j.canlet.2019.02.054.
- [63] C. Chen, X. Ou, J. Wang, W. Guo, y X. Ma, «Radiomics-Based Machine Learning in Differentiation Between Glioblastoma and Metastatic Brain Tumors», *Front. Oncol.*, vol. 9, 2019, doi: 10.3389/fonc.2019.00806.
- [64] S. Bae *et al.*, «Robust performance of deep learning for distinguishing glioblastoma from single brain metastasis using radiomic features: model development and validation», *Sci. Rep.*, vol. 10, n.º 1, Art. n.º 1, jul. 2020, doi: 10.1038/s41598-020-68980-6.
- [65] L. Macyszyn *et al.*, «Imaging patterns predict patient survival and molecular subtype in glioblastoma via machine learning techniques», *Neuro-Oncol.*, vol. 18, n.º 3, pp. 417-425, jun. 2015, doi: 10.1093/neuonc/nov127.
- [66] J. Jeong *et al.*, «Machine-learning based classification of glioblastoma using delta-radiomic features derived from dynamic susceptibility contrast enhanced magnetic resonance images», *Quant. Imaging Med. Surg.*, vol. 9, n.º 7, pp. 1201-1213, jul. 2019, doi: 10.21037/qims.2019.07.01.
- [67] K. Li-Chun Hsieh, C.-Y. Chen, y C.-M. Lo, «Quantitative glioma grading using transformed gray-scale invariant textures of MRI», *Comput. Biol. Med.*, vol. 83, pp. 102-108, abr. 2017, doi: 10.1016/j.compbiomed.2017.02.012.
- [68] N. J. Tustison *et al.*, «Optimal Symmetric Multimodal Templates and Concatenated Random Forests for Supervised Brain Tumor Segmentation (Simplified) with ANTsR», *Neuroinformatics*, vol. 13, n.º 2, pp. 209-225, abr. 2015, doi: 10.1007/s12021-014-9245-2.
- [69] K. Held, E. R. Kops, B. J. Krause, W. M. Wells, R. Kikinis, y H.- Muller-Gartner, «Markov random field segmentation of brain MR images», *IEEE Trans. Med. Imaging*, vol. 16, n.º 6, pp. 878-886, dic. 1997, doi: 10.1109/42.650883.
- [70] C.-H. Lee, M. Schmidt, A. Murtha, A. Bistritz, J. Sander, y R. Greiner, «Segmenting Brain Tumors with Conditional Random Fields and Support Vector Machines», en *Computer Vision for Biomedical Image Applications*, Berlin, Heidelberg, 2005, pp. 469-478. doi: 10.1007/11569541_47.
- [71] D. Zikic *et al.*, «Decision Forests for Tissue-Specific Segmentation of High-Grade Gliomas in Multi-channel MR», en *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2012*, Berlin, Heidelberg, 2012, pp. 369-376. doi: 10.1007/978-3-642-33454-2_46.

- [72] J. Lao *et al.*, «A Deep Learning-Based Radiomics Model for Prediction of Survival in Glioblastoma Multiforme», *Sci. Rep.*, vol. 7, n.º 1, Art. n.º 1, sep. 2017, doi: 10.1038/s41598-017-10649-8.
- [73] G. Wang, W. Li, S. Ourselin, y T. Vercauteren, «Automatic Brain Tumor Segmentation using Cascaded Anisotropic Convolutional Neural Networks», *ArXiv170900382 Cs*, dic. 2017, doi: 10.1007/978-3-319-75238-9_16.
- [74] E. Calabrese, J. E. Villanueva-Meyer, y S. Cha, «A fully automated artificial intelligence method for non-invasive, imaging-based identification of genetic alterations in glioblastomas», *Sci. Rep.*, vol. 10, n.º 1, Art. n.º 1, jul. 2020, doi: 10.1038/s41598-020-68857-8.
- [75] «CPTAC | Office of Cancer Clinical Proteomics Research». <https://proteomics.cancer.gov/programs/cptac> (accedido dic. 17, 2020).
- [76] National Cancer Institute Clinical Proteomic Tumor Analysis Consortium (CPTAC), «Radiology Data from the Clinical Proteomic Tumor Analysis Consortium Glioblastoma Multiforme [CPTAC-GBM] collection». The Cancer Imaging Archive, 2018. doi: 10.7937/K9/TCIA.2018.3RJE41Q1.
- [77] K. Clark *et al.*, «The Cancer Imaging Archive (TCIA): Maintaining and Operating a Public Information Repository», *J. Digit. Imaging*, vol. 26, n.º 6, pp. 1045-1057, dic. 2013, doi: 10.1007/s10278-013-9622-7.
- [78] «General Python FAQ — Python 3.9.1 documentation». <https://docs.python.org/3/faq/general.html#what-is-python> (accedido ene. 22, 2021).
- [79] «Visual Studio Code Frequently Asked Questions». <https://code.visualstudio.com/docs/supporting/faq> (accedido ene. 22, 2021).
- [80] «Python - Visual Studio Marketplace». <https://marketplace.visualstudio.com/items?itemName=ms-python.python> (accedido ene. 22, 2021).
- [81] «PEP 8 -- Style Guide for Python Code», *Python.org*. <https://www.python.org/dev/peps/pep-0008/> (accedido ene. 22, 2021).
- [82] «R: What is R?» <https://www.r-project.org/about.html> (accedido may 26, 2021).
- [83] L. G. Nyúl y J. K. Udupa, «On standardizing the MR image intensity scale», *Magn. Reson. Med.*, vol. 42, n.º 6, pp. 1072-1081, dic. 1999, doi: 10.1002/(sici)1522-2594(199912)42:6<1072::aid-mrm11>3.0.co;2-m.
- [84] L. G. Nyúl, J. K. Udupa, y X. Zhang, «New variants of a method of MRI scale standardization», *IEEE Trans. Med. Imaging*, vol. 19, n.º 2, pp. 143-150, feb. 2000, doi: 10.1109/42.836373.
- [85] K. Rehm, K. Schaper, J. Anderson, R. Woods, S. Stoltzner, y D. Rottenberg, «Putting our heads together: a consensus approach to brain/non-brain segmentation in T1-weighted MR volumes», *NeuroImage*, vol. 22, n.º 3, pp. 1262-1270, jul. 2004, doi: 10.1016/j.neuroimage.2004.03.011.
- [86] «Otsu Thresholding - The Lab Book Pages». <http://www.labbookpages.co.uk/software/imgProc/otsuThreshold.html> (accedido ene. 21, 2021).
- [87] «Extract Brain: Extract Brain Surface (BSE) - MIPAV». [https://mipav.cit.nih.gov/pubwiki/index.php/Extract_Brain:_Extract_Brain_Surface_\(BSE\)](https://mipav.cit.nih.gov/pubwiki/index.php/Extract_Brain:_Extract_Brain_Surface_(BSE)) (accedido ene. 21, 2021).

- [88] W. Benesova, «Segmentation of Brain Tumors from Magnetic Resonance Images using Adaptive Thresholding and Graph Cut Algorithm», 2016. /paper/Segmentation-of-Brain-Tumors-from-Magnetic-Images-Benesova/0cf0cb7439e3de4137382ce40a41ea97a598a885 (accedido ene. 21, 2021).
- [89] «2.1. Gaussian mixture models — scikit-learn 0.24.1 documentation». <https://scikit-learn.org/stable/modules/mixture.html> (accedido ene. 21, 2021).
- [90] L. Vincent, «Morphological grayscale reconstruction in image analysis: applications and efficient algorithms», *IEEE Trans. Image Process.*, vol. 2, n.º 2, pp. 176-201, abr. 1993, doi: 10.1109/83.217222.
- [91] J.-C. Yen, F.-J. Chang, y S. Chang, «A new criterion for automatic multilevel thresholding», *IEEE Trans. Image Process.*, vol. 4, n.º 3, pp. 370-378, mar. 1995, doi: 10.1109/83.366472.
- [92] T. Rohlfing, «Image Similarity and Tissue Overlaps as Surrogates for Image Registration Accuracy: Widely Used but Unreliable», *IEEE Trans. Med. Imaging*, vol. 31, n.º 2, pp. 153-163, feb. 2012, doi: 10.1109/TMI.2011.2163944.

APÉNDICE A. DATOS DE LA EVALUACIÓN

ID	TPR	FPR	FNR	Jaccard
IBSR_01	0,862	0,322	0,138	0,652
IBSR_02	0,852	0,365	0,148	0,624
IBSR_03	0,919	0,194	0,081	0,770
IBSR_04	0,962	0,234	0,038	0,780
IBSR_05	0,879	0,381	0,121	0,636
IBSR_06	0,917	0,374	0,083	0,667
IBSR_07	0,892	0,542	0,108	0,578
IBSR_08	0,951	0,631	0,049	0,583
IBSR_09	0,860	0,354	0,140	0,636
IBSR_10	0,734	0,431	0,266	0,513
IBSR_11	0,266	0,093	0,734	0,244
IBSR_12	0,976	0,591	0,024	0,613
IBSR_13	0,973	0,254	0,027	0,776
IBSR_14	0,954	0,234	0,046	0,774
IBSR_15	0,918	0,212	0,082	0,758
IBSR_16	0,927	0,218	0,073	0,761
IBSR_17	0,921	0,165	0,079	0,790
IBSR_18	0,876	0,160	0,124	0,755



Boxplots de las métricas para la evaluación del algoritmo de extracción del cerebro

APÉNDICE B. RESULTADOS DEL ANÁLISIS ESTÁTICO

☆ PBL-GBM		Passed		Last analysis: 1 minute ago		
Bugs	Vulnerabilities	Hotspots Reviewed	Code Smells	Coverage	Duplications	Lines
A	A	A	A	0.0%	3.3%	1.7k S Python

Resultados del análisis estático del código desarrollado en Python realizado mediante la herramienta SonarQube