

UNIVERSIDAD SIMÓN BOLÍVAR

DECANATO DE ESTUDIOS PROFESIONALES

COORDINACIÓN DE TECNOLOGIA E INGENIERIA ELECTRONICA

GRADO EN INGENIERIA ELECTRONICA

Segmentación de tumores cerebrales en Imágenes de resonancia magnética

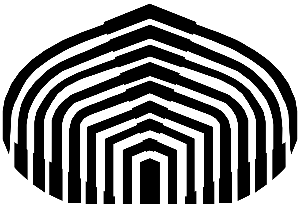
Por:

Juan Diego Ocando Ruiz

PROYECTO DE GRADO

Presentado ante la Ilustre Universidad Simón Bolívar como requisito parcial para optar al título de Ingeniero Electrónico

Sartenejas, Junio de 2019



UNIVERSIDAD SIMÓN BOLÍVAR

DECANATO DE ESTUDIOS PROFESIONALES

COORDINACIÓN DE TECNOLOGIA E INGENIERIA ELECTRONICA

GRADO EN INGENIERIA ELECTRONICA

Segmentación de tumores cerebrales en Imágenes de resonancia magnética

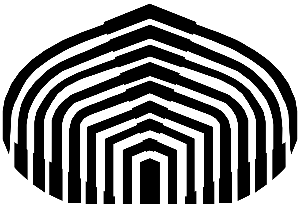
Por:

Juan Diego Ocando Ruiz

Con la asesoría de:

Gerardo Fernández López

Sartenejas, Junio de 2019



UNIVERSIDAD SIMON BOLIVAR

DECANATO DE ESTUDIOS PROFESIONALES

COORDINACION DE INGENIERIA ELECTRONICA

**SEGMENTACION DE TUMORES CEREBRALES EN IMÁGENES DE RESONANCIA MAGNETICA**

PROYECTO DE GRADO

PRESENTADO POR:

Juan Diego Ocando Ruiz, Carnet: 13-10991

**Resumen**

El objetivo de la segmentación de tumores cerebrales en Imágenes de resonancia magnética es el de separar los pixeles o voxeles correspondientes a dichos tumores del resto de la imagen o volumen de la resonancia magnética. Se introduce la justificación y los objetivos del problema. Entonces, se establece una revisión del estado del arte con el fin de introducir al lector sobre el tema y un marco teórico que fundamenta lo que se implementó en los siguientes capítulos. Se presenta un capítulo dedicado al análisis pre-exploratorio de los datos (el cual fue realizado de forma cuantitativa y cualitativa) y de cómo se pueden reducir los datos de forma de equilibrar las diferentes clases a predecir. Esta reducción se basó en el trabajo de Shen (2018). Sin embargo, se tuvieron que realizar ciertas modificaciones con el fin de adaptar dicha reducción a la alta variabilidad y ruido de los datos. La reducción tras las modificaciones recomendadas resulto ser más eficiente en términos de reducción de datos y equilibrio de muestras. Enseguida, se presentó el algoritmo de entrenamiento a utilizar el cual consiste en Stochastic Gradient Descent (SGD) mediante técnicas de validación cruzada. Se establecieron tres casos de prueba: Datos no reducidos, datos reducidos y, tras la reducción de los datos, la aplicación de métodos de Kernel aproximados y análisis de componentes principales para optimizar el desempeño del SGD. Se evaluó cada uno de los casos de prueba y se escogieron los mejores hiperparametros para el Kernel y el SGD.

**Palabras claves:** Segmentación de imágenes, tumores cerebrales, Stochastic Gradient Descent, Validacion cruzada

Página dejada intencionalmente en Blanco

**Invocación de defensa**

Página dejada intencionalmente en Blanco

**Acta de evaluación**

[Chapitre 1 CAPITULO I 15](#_Toc18370467)

[1.1. Justificación 15](#_Toc18370468)

[1.2 Objetivos: 16](#_Toc18370469)

[1.2.1 Objetivo General: 16](#_Toc18370470)

[1.2.3 Objetivos Específicos: 16](#_Toc18370471)

[Estructura del trabajo 16](#_Toc18370472)

[Chapitre 2 18](#_Toc18370473)

[2.1 Tipos de segmentación según el nivel de intervención humana 18](#_Toc18370474)

[2.2 Tipos de segmentación según el grado de supervisión 21](#_Toc18370475)

[2.3 Métodos de segmentación 22](#_Toc18370476)

[2.3.1 Metodos de umbral 22](#_Toc18370477)

[2.3.2 Métodos basados en regiones 23](#_Toc18370478)

[2.3.3 Métodos de clasificación de pixeles 25](#_Toc18370479)

[2.3.4 Métodos de segmentación basados en modelos 27](#_Toc18370480)

[2.4 modelo propuesto 27](#_Toc18370481)

[Chapitre 3 29](#_Toc18370482)

[3.1 Imágenes de resonancia magnética 29](#_Toc18370483)

[3.1.1 Tipos de MRI 29](#_Toc18370484)

[3.2 Cáncer encefálico 30](#_Toc18370485)

[3.3 Análisis de Discriminante de Fischer 31](#_Toc18370486)

[3.4 Support Vector Machine 33](#_Toc18370487)

[3.4.1 Support Vector Machine lineal 33](#_Toc18370488)

[3.4.2 Función de perdida para casos no lineales 35](#_Toc18370489)

[3.5 Gradient descent 36](#_Toc18370490)

[3.6 Stochastic gradient descent 36](#_Toc18370491)

[3.7 Métodos de Kernel 37](#_Toc18370492)

[3.8 K-Means 39](#_Toc18370493)

[3.9 Minibatch K-means 40](#_Toc18370494)

[3.10 Análisis de componentes principales 40](#_Toc18370495)

[3.11 Sobreajuste 41](#_Toc18370496)

[3.12 Validación cruzada 42](#_Toc18370497)

[3.13 Asimetría estadística 43](#_Toc18370498)

[3.14 Curtosis 44](#_Toc18370499)

[Chapitre 4 46](#_Toc18370500)

[4.1 Set de datos utilizado 46](#_Toc18370501)

[4.2 Análisis pre-exploratorio de los datos 47](#_Toc18370502)

[4.2.1 Análisis cuantitativo de los datos 48](#_Toc18370503)

[4.2.2 Análisis cualitativo de los datos 50](#_Toc18370504)

[4.2.3 Dimensionalidad del data set 51](#_Toc18370505)

[4.2.4 Proporcionalidad de tejido tumoral y tejido sano 51](#_Toc18370506)

[4.3 Reducción propuesta 52](#_Toc18370507)

[4.4 Creación de una dataframe de índices de cortes óptimos y procesamiento de la base. 53](#_Toc18370508)

[4.5 Estandarización de datos y eliminación de outliers. 54](#_Toc18370509)

[4.6 Aplicación de agrupación inicial 56](#_Toc18370510)

[4.7 Asignación de subconjuntos a los clúster. 56](#_Toc18370511)

[4.8 Inicialización de SGD 57](#_Toc18370512)

[4.9 Eliminación de clústeres redundantes usando Max-Min cluster Distance 58](#_Toc18370513)

[4.10 Reducción de datos usando FDA 59](#_Toc18370514)

[4.11 Resultados y análisis 62](#_Toc18370515)

[Chapitre 5 CAPITULO IV 67](#_Toc18370516)

[5.1 Modelo propuesto 67](#_Toc18370517)

[5.2 Aplicación de Kernel aproximados y PCA 68](#_Toc18370518)

[5.3 Implementación del esquema de aprendizaje 69](#_Toc18370519)

[5.4 Métricas de clasificación 71](#_Toc18370520)

[5.4.1 Matriz de confusión 71](#_Toc18370521)

[5.4.2 Reporte de clasificación 72](#_Toc18370522)

[5.5 Resultados y análisis 73](#_Toc18370523)

[5.5.1 Análisis de rendimiento de SGD para los datos no reducidos. 75](#_Toc18370524)

[5.5.2 Análisis de desempeño de SGD para los datos reducidos 85](#_Toc18370525)

[5.5.3 Análisis de desempeño para los métodos de Kernel 93](#_Toc18370526)

[5.5.4 Recapitulación 100](#_Toc18370527)

[Chapitre 6 102](#_Toc18370528)

[Chapitre 7 106](#_Toc18370529)

[Apéndice A 111](#_Toc18370530)

[Python 111](#_Toc18370531)

[Numpy 111](#_Toc18370532)

[Pandas 111](#_Toc18370533)

[Sci-kit Learn 111](#_Toc18370534)

[Matplob lib 112](#_Toc18370535)

[Herramientas de hardware 112](#_Toc18370536)

[Apéndice B 113](#_Toc18370537)

[Análisis pre-exploratorio detallado de cada paciente 113](#_Toc18370538)

[Apéndice C 114](#_Toc18370539)

[Código comentado 114](#_Toc18370540)

[Ilustración 2.1: Un tumor en modalidad T1 y modalidad T2 y la diferenciación entre tejido activo, necrótico y edema [1] 18](#_Toc18370541)

[Ilustración 2.2 Segmentación de cuatro expertos diferentes del mismo tumor [1, 3]. Obsérvese que existe una intervariabilidad de los resultados 19](#_Toc18370542)

[Ilustración 2.3 Segmentación automática usando métodos de atlas (Datos de cerebro sano) [15] 21](#_Toc18370543)

[Ilustración 2.4 Diferencias entre aplicar métodos de umbral local y métodos de umbral global [20] 23](#_Toc18370544)

[Ilustración 2.5 Ejemplo del algoritmo de crecimiento de regiones [21] 24](#_Toc18370545)

[Ilustración 2.6 A la izquierda: El gradiente de una imagen. A la derecha: el equivalente topográfico de dicha imagen. [23] 24](#_Toc18370546)

[Ilustración 2.7 Ejemplo de una red neuronal artificial 26](#_Toc18370547)

[Ilustración 2.8: Arquitectura de Red neuronal propuesta por el equipo de BioMedIA [25] 26](#_Toc18370548)

[Ilustración 2.9 Volumen deformable en 3D del corazón de un canino [27] 27](#_Toc18370549)

[Ilustración 2.10 Esquema del proyecto. En naranja los temas que forman parte del capitulo 4 y en Azul aquellos que forman parte del capitulo 5. 28](#_Toc18370550)

[Ilustración 3.1 Diferentes tipos de relajación de un tejido especifico 30](#_Toc18370551)

[Ilustración 3.2 Progresión de un glioblastoma multiforme tras un mes sin tratamiento. [38] 31](#_Toc18370552)

[Ilustración 3.3 Proyección de un conjunto de datos con respecto a dos líneas diferentes. Se puede observar que la recta w1 muestra una mejor separación entre las dos clases de la imagen que la recta w2 32](#_Toc18370553)

[Ilustración 3.4 SVM en un dataset de 2D 34](#_Toc18370554)

[Ilustración 3.5 Demostración de un subgradiente. El punto X1 tiene una sola gradiente, sin embargo, el punto X2 tiene diferentes subgradientes. 37](#_Toc18370555)

[Ilustración 3.6 Algunos de los Kernels más utilizados [39] 39](#_Toc18370556)

[Ilustración 3.7 Evolución de K-means 39](#_Toc18370557)

[Ilustración 3.8 Los dos ejes de componentes principales en la foto de un pescado 41](#_Toc18370558)

[Ilustración 3.9. Ejemplo de un modelo de aprendizaje que representa sobreajuste 41](#_Toc18370559)

[Ilustración 3.10. Ejemplo de una validacion cruzada. 43](#_Toc18370560)

[Ilustración 3.11 Diferentes valores de asimetria estadistica [42] 43](#_Toc18370561)

[Ilustración 3.12 Variabilidad de una señal según su curtosis. 45](#_Toc18370562)

[Ilustración 4.1 Diagrama de dispersión en 3D para un solo corte desde diferentes ángulos. Violeta: Tejido cerebral sano. 50](#_Toc18370563)

[Ilustración 4.3 Diagrama de entrada y salida de los diferentes bloques funcionales de código. 52](#_Toc18370564)

[Ilustración 4.4 : Tres cortes principales de una tomografía 54](#_Toc18370565)

[Ilustración 4.5 Se muestra una iteración de SGD para el mismo dominio sin estandarizar y tras ser estandarizado. Se puede apreciar cómo para el mismo punto y misma gradiente, resulta más óptimo estandarizar las dimensiones de los datos. 55](#_Toc18370566)

[Ilustración 4.6 : Eliminación de clusters redundantes.[37] 59](#_Toc18370567)

[Ilustración 4.7 : Eliminación de datos redundantes en un solo clúster unitario. [37] 60](#_Toc18370568)

[Ilustración 4.8 : Modificación del algoritmo FIFDR propuesto originalmente por el autor. A la izquierda el esquema original, a la derecha la parte modificada. En naranja la etapa que ha sido modificada. 62](#_Toc18370569)

[Ilustración 4.9 Comportamiento del FDR en un clúster aleatorio y sus diferentes parámetros. 65](#_Toc18370570)

[Ilustración 5.1 Esquema propuesto de modelo de aprendizaje 68](#_Toc18370571)

[Ilustración 5.2 ilustracion de diferentes metricas de reporte de aprendizaje 72](#_Toc18370572)

[Ilustración 5.3 Esquema propuesto de evaluacion para un solo puntaje definido. 75](#_Toc18370573)

[Ilustración 5.4 Validación visual con el paciente 150 y datos no reducidos. 82](#_Toc18370574)

[Ilustración 5.5 Validación visual con el paciente 185 y datos no reducidos. 83](#_Toc18370575)

[Ilustración 5.6 Validación visual con el paciente 186 y datos no reducidos. 84](#_Toc18370576)

[Ilustración 5.7 Validación visual con el paciente 205 y datos no reducidos. 84](#_Toc18370577)

[Ilustración 5.8 Validación visual paciente 150, en caso de datos reducidos 90](#_Toc18370578)

[Ilustración 5.9 Validacion visual paciente 185, en caso de datos reducidos 91](#_Toc18370579)

[Ilustración 5.10 Validación visual paciente 186, en caso de datos reducidos 92](#_Toc18370580)

[Ilustración 5.11 Validación visual 205, en caso de datos reducidos 93](#_Toc18370581)

[Ilustración 5.12 Validación visual del paciente 150, para el caso de datos reducidos y modificados mediante métodos de Kernel aproximados 97](#_Toc18370582)

[Ilustración 5.13 Validación visual del paciente 185, para el caso de datos reducidos y modificados mediante métodos de Kernel aproximados 98](#_Toc18370583)

[Ilustración 5.14 Validación visual del paciente 186, para el caso de datos reducidos y modificados mediante métodos de Kernel aproximados 99](#_Toc18370584)

[Ilustración 5.15 Validacion del paciente 205, para el caso de datos reducidos y modificados mediante metodos de Kernel aproximados. 100](#_Toc18370585)

[Tabla 4.1 Analisis estadistico de las muestras 49](#_Toc18361945)

[Tabla 4.2 Proporcion entre tejidos tumorales y tejidos sanos en promedio por cada paciente. 52](#_Toc18361946)

[Tabla 4.3 Encabezado del dataframe de coordenadas de cortes optimos. 53](#_Toc18361947)

[Tabla 4.4 Dataframe de datos de voxeles para entrenar el algoritmo en amarillo voxeles del corte coronal; en azul, corte sagital y en verde, corte horizontal. Cada celda representa, en realidad, todas las entradas representadas por corte optimizado de paciente. 54](#_Toc18361948)

[Tabla 4.5 : Reducción de datos a lo largo del proceso propuesto aplicando un Threshold de 0.1 en la eliminación de clústeres redundantes. 63](#_Toc18361949)

[Tabla 4.6 Reducción de datos a lo largo del proceso propuesto por Shen aplicado sobre el mismo dataset. 63](#_Toc18361950)

[Tabla 4.7Proporcion entre clases para cada cluster. Los clusters resaltados representan clusteres unitarios y seran candidatos para el algoritmo MAX-MIN Dist. 64](#_Toc18361951)

[Tabla 4.8 Clusteres redundantes eliminados tras la aplicación del algorimo MAX-Min distancias.. 64](#_Toc18361952)

[Tabla 5.1 Matriz de confusion 71](#_Toc18361953)

[Tabla 5.2 Promedio y error de los resultados de precisión de la validación cruzada para los diferentes parámetros establecidos en la búsqueda de rejilla con el dataset no reducido. 76](#_Toc18361954)

[Tabla 5.3 Promedio y error de los resultados de sensibilidad de la validación cruzada para los diferentes parámetros establecidos en la búsqueda de rejilla con el dataset no reducido. 77](#_Toc18361955)

[Tabla 5.4 Promedio y error de los resultados de f1score de la validación cruzada para los diferentes parámetros establecidos en la búsqueda de rejilla con el dataset no reducido. 78](#_Toc18361956)

[Tabla 5.5 Reporte de clasificación de los resultados de precisión para el dataset no reducido. 79](#_Toc18361957)

[Tabla 5.6 Matriz de confusión de los resultados de precisión para el dataset no reducido. 79](#_Toc18361958)

[Tabla 5.7 Reporte de clasificación de los resultados de Sensibilidad para el dataset no reducido. 80](#_Toc18361959)

[Tabla 5.8 Matriz de confusión de los resultados de sensibilidad para el dataset no reducido. 80](#_Toc18361960)

[Tabla 5.9 Reporte de clasificación de los resultados del puntaje F1 para el dataset no reducido. 80](#_Toc18361961)

[Tabla 5.10 Matriz de confusión de los resultados del puntaje F1 para el dataset no reducido. 81](#_Toc18361962)

[Tabla 5.11 Tiempo de desarrolo de datos reducidos vs Datos No reducidos 85](#_Toc18361963)

[Tabla 5.12 Promedio y error de los resultados de precisión de la validación cruzada para los diferentes parámetros establecidos en la búsqueda de rejilla con el dataset reducido 86](#_Toc18361964)

[Tabla 5.13 Promedio y error de los resultados de sensibilidad de la validación cruzada para los diferentes parámetros establecidos en la búsqueda de rejilla con el dataset reducido 87](#_Toc18361965)

[Tabla 5.14 Promedio y error de los resultados del puntaje F1 de la validación cruzada para los diferentes parámetros establecidos en la búsqueda de rejilla con el dataset reducido 87](#_Toc18361966)

[Tabla 5.15 Reporte de clasificación de los resultados de precision para el dataset reducido 88](#_Toc18361967)

[Tabla 5.16 Matriz de confusion de los resultados de precision para el dataset reducido 88](#_Toc18361968)

[Tabla 5.17 Reporte de clasificación de los resultados de sensibilidad para el dataset reducido 89](#_Toc18361969)

[Tabla 5.18 Matriz de confusion de los resultados de sensibilidad para el dataset reducido 89](#_Toc18361970)

[Tabla 5.19 Reporte de clasificación de los resultados del puntaje F1 para el dataset reducido 89](#_Toc18361971)

[Tabla 5.20 Matriz de confusion de los resultados del puntaje F1 para el dataset reducido 89](#_Toc18361972)

[Tabla 5.21 Fase de desarrolo del caso de datos reducidos y modificados mediante metodos de Kernel aproximados. 94](#_Toc18361973)

[Tabla 5.22 Reporte de clasificacion para el modelo de datos reducidos con aplicacion de metodos de Kernal aproximado, en la fase evaluación cuantitativa. 95](#_Toc18361974)

[Tabla 5.23 Matriz de confusion para el modelo de datos reducidos con aplicacion de metodos de Kernal aproximado, en la fase evaluación cuantitativa. 95](#_Toc18361975)

[Tabla 5.24 Matriz de confusion para el paciente 150, en el caso de datos reducidos y modificados mediante métodos 96](#_Toc18361976)

[Tabla 5.25 Reporte de clasificacion para el paciente 150, en el caso de datos reducidos y modificados mediante métodos de Kernel aproximados 96](#_Toc18361977)

[Tabla 5.26 Matriz de confusion para el paciente 185, en el caso de datos reducidos y modificados mediante métodos 97](#_Toc18361978)

[Tabla 5.27 Reporte de clasificacion para el paciente 185, en el caso de datos reducidos y modificados mediante métodos 97](#_Toc18361979)

[Tabla 5.28 Matriz de confusion para el paciente 186, en el caso de datos reducidos y modificados mediante métodos 98](#_Toc18361980)

[Tabla 5.29 Reporte de clasificacion para el paciente 186, en el caso de datos reducidos y modificados mediante métodos 99](#_Toc18361981)

[Tabla 5.31 Matriz de confusion para el paciente 205, en el caso de datos reducidos y modificados mediante métodos 100](#_Toc18361982)

[Tabla 5.30 Reporte de clasificacion para el paciente 205, en el caso de datos reducidos y modificados mediante métodos 100](#_Toc18361983)

Glosario

**SVM:** Support vector machines.

**PCA:** Análisis de componentes principales (Principal Components Analysis)

**IRM:** Imagen de resonancia magnética

**T1:** Tiempo de relajación T1

**T2:** Tiempo de relajación T2

**T1C:** Tiempo de relajación T1 con agente de contraste

**FLAIR:** Fluid Attenuated Inversion Recovery

**FDA:** análisis discriminante de Fischer (Fischer Discriminant Analysis)

**FDR:** Radio de discriminante de Fischer (Fischer Discriminant Radio)

**SGD:** Stochastic Gradient Descent

**CV**: cross-validation (validación cruzada)

**RBF**: Radial basis function (función de base radial)

**FIFDR**: Fast iteration Fischer Discriminatory Ratio (radio discriminante de Fischer en iteraciones rapidas)

# CAPITULO I

## 1.1. Justificación

El cáncer es una de las principales causas de muerte en Venezuela y en el mundo. Algunos tipos de cáncer pueden generar masas anormales de tejido conocidos como tumores o neoplasias. Entre estos, el cáncer de cerebro, a pesar de encontrarse entre los tipos menos comunes, es de los que causa un mayor impacto debido al mal pronóstico que este suele tener. Los tumores que se pueden encontrar en el encéfalo son diagnosticados mediante exámenes imageneológicos como la tomografía axial computarizada (TAC) o la resonancia magnética (RM). Las imágenes de resonancia magnética (IRM) son las preferidas para el diagnóstico de tumores cerebrales, ya que estas ofrecen imágenes más detalladas que las obtenidas usando la TAC. Las imágenes obtenidas en TAC son preferidas en casos donde es necesario analizar como el tumor afecta el cráneo, debido a que dichas imágenes muestran con mayor detalle los huesos.

Una vez obtenida la imagen usando los exámenes imagenológicos, es necesario, realizar una segmentación del tumor, es decir, separar los diferentes tejidos del tumor (solido o activo, edema, necrosis) del tejido normal del cerebro: materia gris, materia blanca y fluido cerebroespinal en la imagen [1] La detección temprana y la localización del tumor junto a una estadificación precisa del cáncer influyen en el tratamiento de forma importante y afectan el resultado de la salud del paciente. Por lo tanto, el éxito en el tratamiento del paciente depende en bastante grado de una delimitación y segmentación precisa del tumor o área de interés en las imágenes. Sin embargo, en la segmentación de tumores cerebrales, se ha observado que existen, en promedio, variaciones de 28 ± 12 % entre operadores al ejecutar la misma tarea y variaciones de 20 ± 15% al momento en el que el mismo operador repite la misma tarea en intervalos de un mes. [2]

Es por ello por lo que, usando algoritmos de procesamiento de imágenes y de aprendizaje automático, se puede crear una herramienta capaz de realizar la segmentación del tumor en IRM y ser de utilidad para que los especialistas puedan minimizar las ambigüedades que pueden surgir al realizar el proceso de segmentación manual, y que, a partir de los resultados sugeridos por los algoritmos, se pueda lograr una estadificación precisa del cáncer, escoger un tratamiento adecuado y mejorar el pronóstico del paciente.

## 1.2 Objetivos:

### 1.2.1 Objetivo General:

Desarrollar algoritmos para segmentar tumores cerebrales en imágenes de resonancia magnética.

### 1.2.3 Objetivos Específicos:

* Analizar las características de imágenes de resonancia magnéticas del encéfalo y de sus regiones de interés.
* Definir algoritmos de preprocesamiento para optimizar los procesos de segmentación de tumores cerebrales.
* Distinguir la región de interés del encéfalo de forma automática usando algoritmos de procesamientos de imágenes.
* Interpretar los resultados sugeridos por los distintos algoritmos planteados.

## Estructura del trabajo

Este trabajo se encuentra dividido en seis capítulos. El **capítulo 1,** se trata de la introducción, la justificación, la descripción del problema y la exposición de los objetivos generales y específicos del proyecto.

**El capítulo 2** explica el estado del arte de la segmentación de tumores cerebrales en imágenes de resonancia magnética, explicando el problema de forma general, explicando la clasificación de los diferentes métodos, algunos de los algoritmos más utilizados y sus ventajas y desventajas. Adicionalmente, al final de capitulo se ofrece una breve introducción de cuál ha sido el método seleccionado para el trabajo y una breve justificación de por qué se decidió utilizarlo.

El marco teórico de los términos y algoritmos utilizados a lo largo del proyecto forma parte del **capítulo 3**. Se dan a entender conceptos importantes sobre el cáncer, los gliomas y las diferentes modalidades de imágenes de resonancia magnética utilizadas; así como las formulas detrás de los diferentes algoritmos establecidos y conceptos estadísticos utilizados para entender los datos con mayor profundidad.

Los datos y su procesamiento son un elemento clave del proyecto. Por lo tanto, el **capítulo 4** es dedicado exclusivamente a dicho tema. Se expone y se analiza el conjunto de datos utilizado, al mismo tiempo que se detalla cómo ha sido el proceso de procesamiento de dichos datos y como estos han sido puestos en un formato establecido.

El capítulo 5, describe el proceso de aprendizaje para lograr la segmentación de tumores cerebrales, a partir de los conjuntos de datos obtenidos al final del capítulo anterior. Adicionalmente, en este capítulo se incluye la evaluación y análisis de dichos resultados.

Finalmente, el capítulo 6 es la conclusión de dichos análisis y donde, finalmente, se establecen recomendaciones para posibles futuros trabajos o se proponen mejoras respecto al trabajo actual.

# 

REVISIÓN DEL ESTADO DEL ARTE

El presente capítulo se hará un resumen del estado actual de las investigaciones de la segmentación de tumores en imágenes de resonancia magnética. Así como una explicación de la clasificación que existe de los diferentes métodos. Es importante resaltar que esta investigación está basada en el trabajo que Nelly Gordillo realizo en el 2013. [1] Para aquellos lectores con conocimientos sobre el tema, se recomienda avanzar directamente a la **sección 2.4**, donde se explica brevemente el modelo que se ejecutó durante este proyecto.

El objetivo principal de la segmentación de imágenes consiste en extraer los pixeles pertenecientes a una región de interés del resto de la imagen. En el caso de tumores cerebrales en IRM, se trata de separar los diferentes tejidos de un tumor cerebral (Edema, tumor activo y tumor necrótico) del resto del tejido sano del cerebro (Materia blanca, materia gris y líquido cefalorraquídeo). [1]

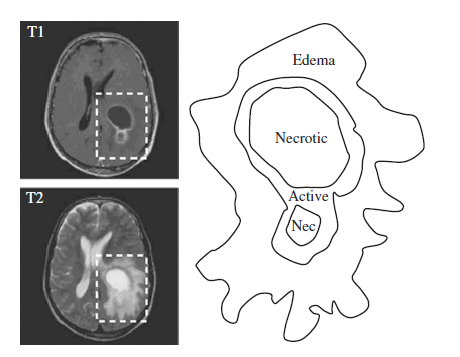


Ilustración 2.1: Un tumor en modalidad T1 y modalidad T2 y la diferenciación entre tejido activo, necrótico y edema [1]

## 2.1 Tipos de segmentación según el nivel de intervención humana

La segmentación de tumores cerebrales se divide en tres principales categorías de acuerdo al grado de interacción humana que se requiere para ejecutar la tarea. Se describen como clasificaciones la segmentación manual, semiautomática y automática.

**La segmentación manual** de tumores en imágenes es el simple hecho de trazar manualmente los límites del tumor y las diferentes estructuras de interés, o pintar la región de las estructuras anatómicas según diferentes etiquetas. Normalmente, intervienen especialistas en el área como radiólogos, anatomistas o técnicos entrenados para ejercer este tipo de tarea con la ayuda de herramientas de software que permiten analizar las imágenes mediante una interfaz gráfica y, de este modo, facilitar el trazado de los límites en las diferentes regiones de interés. Esta actividad requiere de bastante tiempo y entrenamiento, dado que los escáneres de IRM generan múltiples imágenes de dos dimensiones y el experto tiene que pasar por el conjunto de imágenes corte tras corte para delinear cuidadosamente la región de interés. Por lo tanto, la segmentación manual se basa típicamente en una sola imagen con agentes de contrastes aplicados de forma intravenosa al paciente.

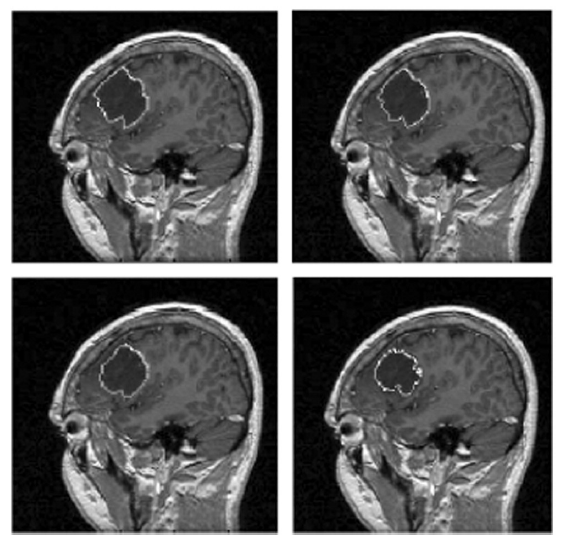


Ilustración 2.2 Segmentación de cuatro expertos diferentes del mismo tumor [1, 3]. Obsérvese que existe una intervariabilidad de los resultados

Como se ha explicado en la justificación de este trabajo, la segmentación de tumores cerebrales presenta un gran grado de dependencia del experto que ejecuta la tarea, obteniéndose importantes valores de inter e intra variabilidad. A pesar de esto, los resultados manuales de segmentaciones son utilizados como verdad fundamental para los algoritmos automáticos y semiautomáticos como explica Wu [10].

**Los métodos semiautomáticos** de segmentación requieren de la intervención de un operador humano para ejecutar alguna de las siguientes tareas, estas fueron definidas según el estudio de Foo [11] y son citadas en el trabajo de Gordillo [1]:

* **Inicialización**: las etapas de este proceso comprenden la entrada de los parámetros o argumentos para aplicar los algoritmos, la aplicación de algunas técnicas de preprocesamiento en los datos de la imágenes para mejorar la calidad, la evaluación de los datos de la imagen para mejorar las decisiones de parámetros o la elección de una región inicial para ser corregida por el algoritmo.
* **Intervención o respuesta al feedback:** corregir el resultado continuamente hasta alcanzar un resultado deseado, respondiendo al feedback aportado por el algoritmo en cada iteración hasta detener el procedimiento en el momento en el que el resultado se considere aceptable.



* **Evaluación:** Evaluar si los resultados son deseados y decidir de repetir un procedimiento de inicialización para obtener mejores resultados o aceptar los resultados finales.

El componente principal de estos métodos es la parte computacional, sin embargo, al requerir la intervención humana, los resultados pueden variar según la aplicación de cada usuario.



**La segmentación automática** no requiere de ninguna intervención humana. Los métodos completamente automatizados utilizan en su mayoría inteligencia artificial o aplicación de conocimientos previos en sus algoritmos. Es común el uso de *soft-computing* (resolución de problemas donde no se puede alcanzar una respuesta exacta. Esto representa un tema de bastante estudio en el campo del aprendizaje de maquina ya que representa un problema que los humanos pueden hacer efectivamente tras cierto entrenamiento y experiencia. Sin embargo, realizar estos métodos con una gran precisión representa gran dificultad, debido a que los humanos deben usar un procesamiento visual de alto nivel y deben incorporar conocimientos propios del dominio radiológico para ejecutar esta tarea. [1]





Ilustración 2.3 Segmentación automática usando métodos de atlas (Datos de cerebro sano) [15]

## 2.2 Tipos de segmentación según el grado de supervisión

Existen dos formas principales para alcanzar el objetivo de la segmentación: los métodos supervisados y no supervisados. **Los métodos no supervisados** son aquellos donde las etiquetas de los pixeles/voxeles de una imagen son desconocidas, por lo tanto, se aportan los datos sin los resultados previos de un humano ni subconjuntos de datos. En general, se somete a todo el conjunto de datos a métodos de clúster para conseguir las clasificaciones.

Las tres desventajas principales de los métodos no supervisados son señaladas en el trabajo de Schmidt [12]: Se requiere especificar con anterioridad el número de regiones a utilizar; los tumores pueden estar divididos en múltiples regiones y estos no suelen tener un nivel definido de intensidad o de barreras de textura. Algunas técnicas para superar estas desventajas es reducir las homogeneidades y aplicar segmentación intracraneal, es decir, no considerar el cráneo en la imagen.

Los **métodos supervisados** se diferencian de los no supervisados por el uso explícito de datos etiquetados (verdad fundamental) para el entrenamiento de los modelos. Estos requieren de una fase de entrenamiento para poder clasificar un conjunto específico de características en etiquetas y una fase de prueba para confirmar el funcionamiento del modelo. En estos modelos, una buena elección y definición de la estructura de datos es crucial ya que diferentes estructuras de datos pueden afectar el tiempo de entrenamiento y los resultados de la segmentación.

La mayor ventaja en los métodos supervisados radica en que se pueden realizar diferentes tareas simplemente cambiando el conjunto de datos. Por lo tanto, son más efectivos y versátiles. Sin embargo, la mayor desventaja es que en muchos escenarios se necesita entrenamiento para casos muy específicos y la variabilidad en la verdad fundamental (ya que esta se basa en segmentaciones manuales) puede ser otra desventaja.

## 2.3 Métodos de segmentación

Diferentes autores [1, 14] definen los métodos de segmentación en 4 categorías principales: Métodos de umbral, Métodos basados en umbral, métodos basados en regiones, métodos basados en clasificación de pixeles y métodos basados en modelos.

### 2.3.1 Metodos de umbral

Los objetos son comparados según su intensidad dependiendo de uno o más umbrales, basándose en los histogramas de las imágenes, siempre que estos puedan ser divididos en patrones bimodales. Sin embargo, en caso de tener más de dos tipos de regiones que corresponden a diferentes objetos, se debe aplicar un umbral local.

Los **métodos de umbral global** consisten en clasificar binariamente los pixeles cuyas intensidades son menores a un umbral especifico denominado , (tejido no tumoral) y aquellos con intensidades superiores a en otra categoría (tejido tumoral).

Se puede definir mediante la siguiente funcion de clasificación donde e son las coordenadas de un pixel y es la función intensidad.

El mayor problema de estos métodos es el de no poder definir precisamente un umbral para los bordes de los tumores, por lo tanto, existe un gran riesgo de obtener tumores con volúmenes mayores o menores al real.

Los **métodos de umbral local** [16] consisten en aplicar un umbral adaptativamente para cada pixel. Es decir, se analiza cada pixel de forma individual y sus pixeles vecinos con el fin de establecer un umbral específico para dicho pixel.

Estas técnicas permiten detectar el valor de un pixel cuando este no puede ser definido fácilmente mediante el histograma de la imagen completa. Generalmente los valores de estos umbrales son determinados a partir de conocimientos previos de la imagen. Sin embargo, estos métodos se consideran complejos en tiempo desde un punto de vista computacional y no se aconsejan para aplicaciones en tiempo real.

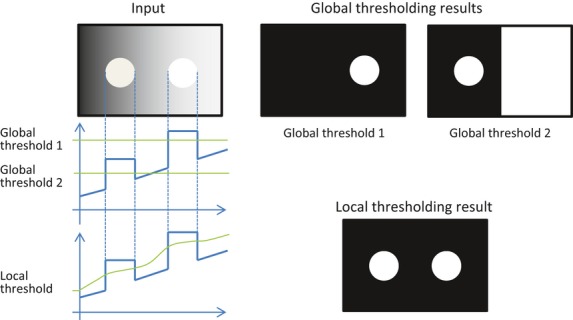


Ilustración 2.4 Diferencias entre aplicar métodos de umbral local y métodos de umbral global [20]

### 2.3.2 Métodos basados en regiones

Los métodos basados en regiones [1] examinan un conjunto de pixeles en una imagen y forman regiones separadas que incluyen pixeles cercanos con valores de homogeneidad basando en un criterio de similitud ya establecido.

El esquema general consiste en tener una imagen dividida en regiones denominadas donde . Se tiene , se cumple la condición . Se define un predicado lógico que contiene un conjunto de reglas establecido que se aplica a todos los pixeles dentro de la misma región que cumple las siguientes propiedades

Entre los métodos más utilizados existen aquellos de crecimiento regiones [21] el cual consiste en eliminar en una región de la imagen basada en un criterio específico. Se selecciona una serie de pixeles inicial, se analizan iterativamente los pixeles vecinos de cada uno y si cumplen con una condición establecida (como, por ejemplo, cierta similitud en la intensidad con el pixel original) se considera como parte de la misma región que el pixel inicial.

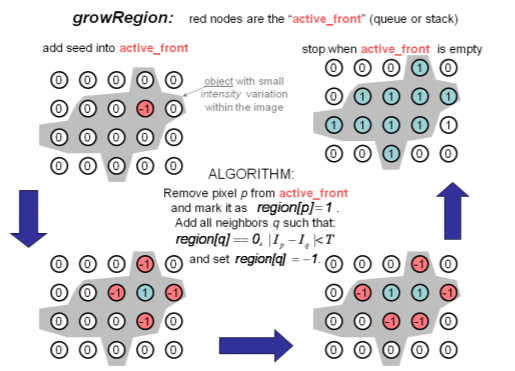


Ilustración 2.5 Ejemplo del algoritmo de crecimiento de regiones [21]

El método de **transformación divisoria** (o *watershed* en inglés) [23, 24] consiste en utilizar una imagen como un relieve fotográfico; donde se interpreta el nivel de intensidad como la altura de la imagen. Cada pixel se considera como una gota de agua que fluye hasta converger hasta un mínimo local (que topográficamente se puede considerar como una cuenca hidrográfica). Las diferentes regiones que se obtienen corresponden a los límites de las cuencas hidrográficas halladas.

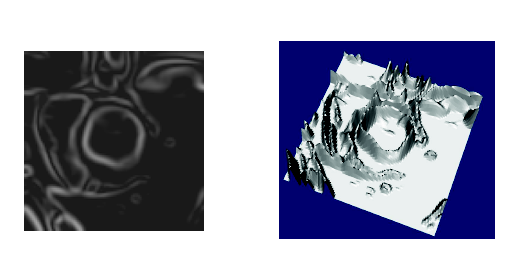


Ilustración 2.6 A la izquierda: El gradiente de una imagen. A la derecha: el equivalente topográfico de dicha imagen. [23]

### 2.3.3 Métodos de clasificación de pixeles

En este caso se toman valores específicos de cada pixel y se trasladan al espacio de características. Se puede aplicar para imágenes de un solo canal o en imágenes multimodales donde se crea un espacio de características multifactorial. Estos métodos están restringidos al uso de métodos de agrupamiento de forma supervisada o no supervisada.

Por ejemplo, entre los métodos no supervisados se utilizan: *Fuzzy C-means*, k-means y métodos estadísticos como las cadenas aleatorias de Markov. Por otro lado, como métodos supervisados, existen las redes neuronales artificiales, los métodos bayesianos y otros.

**Fuzzy C-means** es utilizado porque, en muchas situaciones, no es sencillo determinar si un pixel pertenece a una región específica o no, ya que sus características no permiten determinar una transición evidente entre los límites de diferentes regiones. Entonces, se introducen los conceptos difusos y cada muestra pueda pertenecer a más de un clúster o grupo al mismo tiempo. Como este algoritmo es iterativo, se considera que consume mucho tiempo en el caso de tener una gran cantidad de muestras, para lograr a converger a un resultado aceptable.

**Los campos aleatorios de Markov** (*Markov Random Fields* en inglés)permiten considerar las condiciones espaciales del pixel dentro del algoritmo, e integrar dicha información al proceso de agrupación. Es decir, lo que no es posible lograr utilizando algoritmos como Fuzzy C-means o K-means. Este algoritmo tiene altos niveles de precisión, al determinar que si un pixel tiene una fuerte tendencia a pertenecer a una clase (sea tejido tumoral o no tumoral), sus pixeles vecinos tienen una tendencia importante a compartir la misma clase que dicho pixel y de este modo representar dependencias complejas entre las diferentes muestras. [1] La principal desventaja: el alto costo computacional. Sin embargo, existen alternativas como los SVM.

Las **Redes Neuronales Artificiales** (ANN) [26] funcionan a través de una serie de nodos o neuronas, que están compuestas por una serie de entradas, una serie de capas ocultas y una capa de salida. Para cada entrada, se aplica una serie de operaciones matemáticas interconectadas entre los diferentes nodos para obtener una clasificación en los nodos de salida.

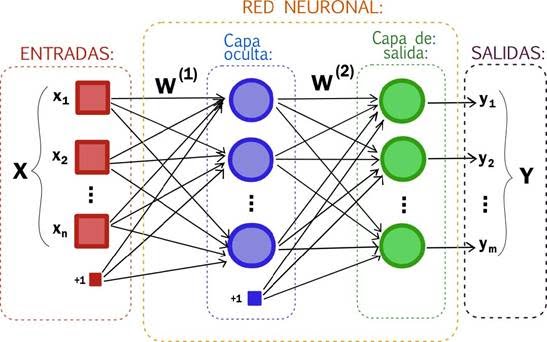


Ilustración 2.7 Ejemplo de una red neuronal artificial

Entre las implementaciones más destacas de algoritmos de redes neuronales artificiales se encuentra la arquitectura diseñada por el equipo de BioMedIA compuesto por 11 estudiantes de PhD del Imperial College London y ganadores del reto BraTS 2017, que cuenta con dos arquitecturas en paralelo de redes neuronales convolucionales.. [25]

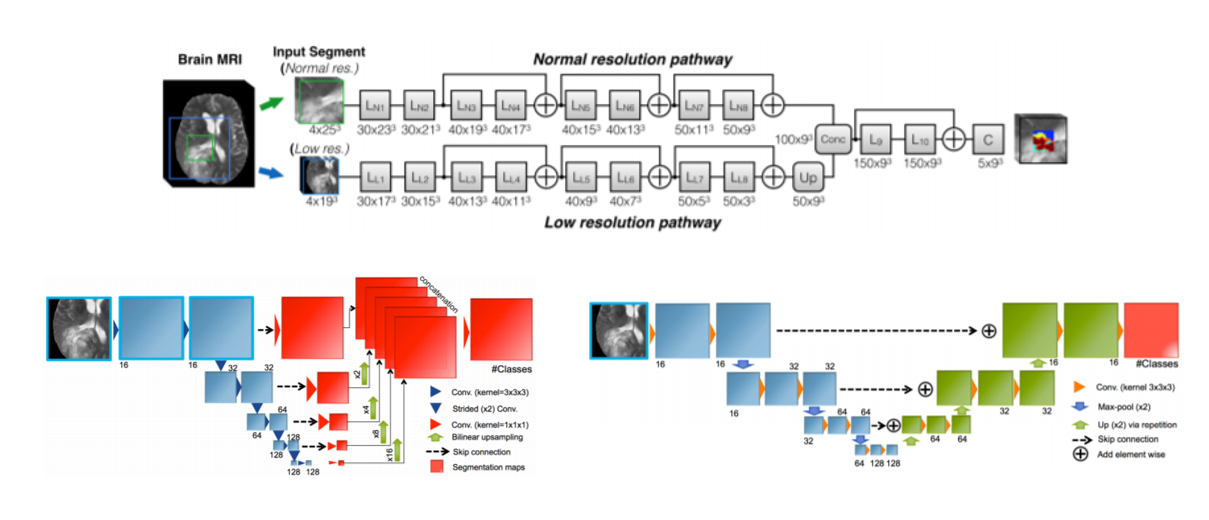


Ilustración 2.8: Arquitectura de Red neuronal propuesta por el equipo de BioMedIA [25]

Como se puede observar en la imagen anterior, el tamaño de las redes neuronales que pueden ejecutar esta tarea de forma eficiente es muy grande y, por lo tanto, su tiempo de entrenamiento es inaceptable, mientras se requiere también de un gran número de imágenes para poder entrenar la red y de una infraestructura para aplicarlo. Sin embargo, con la reducción de precios de tarjetas gráficas avanzadas, estos algoritmos son usados cada vez con más frecuencia.

### 2.3.4 Métodos de segmentación basados en modelos

A diferencia de los métodos anteriores que son aplicados normalmente en imágenes de 2D, los métodos basados en modelos son normalmente aplicados en un contexto volumétrico. Se crea un modelo continuo y anatómico para una estructura específica mientras se incorporan conocimientos previos de los objetos como forma, ubicación y localización. Algunos modelos incorporan información estadística a partir de una población de conjuntos de datos de entrenamiento. Segmentar estructuras a partir de imágenes previas es un proceso complejo debido al gran tamaño de los conjuntos de datos y la variabilidad y complejidad anatómica de las formas de los diferentes gliomas. Los modelos deformables comprenden la generación de una interfaz de propagación (en 2D se define como una curva cerrada y en 3D como una superficie cerrada) cuya forma varía dependiendo de parámetros independientes, sean estos locales o globales.

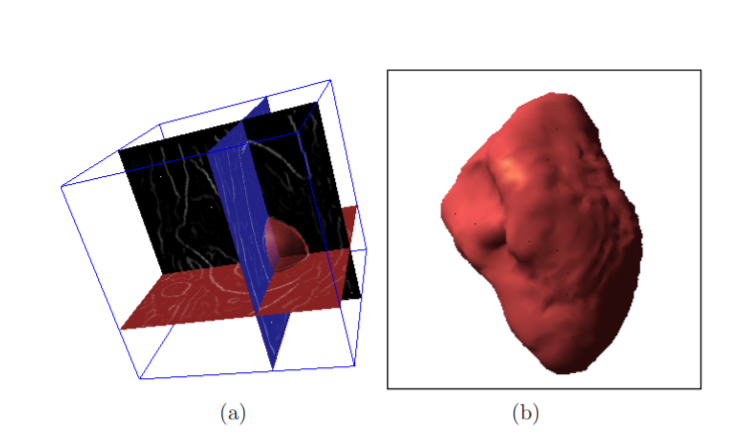


Ilustración 2.9 Volumen deformable en 3D del corazón de un canino [27]

## 2.4 modelo propuesto

Considerando que la mayoría de los modelos expuestos anteriormente son relativamente lentos para ejecutar un proceso de aprendizaje, que requieren de requerimientos de hardware con una amplia capacidad y que el volumen de datos a analizar para este proyecto es bastante voluminoso (ver sección 4.2.3). Se propone un esquema basado en algoritmos estocásticos que puedan superar esta problemática.

Este esquema se ejecutara de una forma lineal: se propone una reducción y pre procesamiento de datos con el fin de optimizar el tiempo de aprendizaje utilizado. La reducción de datos está basada en un modelo de agrupamiento que utiliza Mini Batch K-means y el discrimante lineal de Fischer, con el fin de eliminar todos los datos redundantes del espacio de características y, sin embargo, permanecer con datos que tengan suficiente relevancia para ejecutar la clasificación.

Finalmente, se ejecutara un modelo de aprendizaje el cual constara de las etapas de entrenamiento, validación y evaluación. Este está basado principalmente en el uso de métodos aproximados de Kernel y la aplicación de Stochastic Gradient Descent mediante la utilización de validación cruzada. En este caso, estos algoritmos forman parte de los métodos de clasificación de pixeles, debido a que se hace un mapeo de los pixeles en el espacio de características.

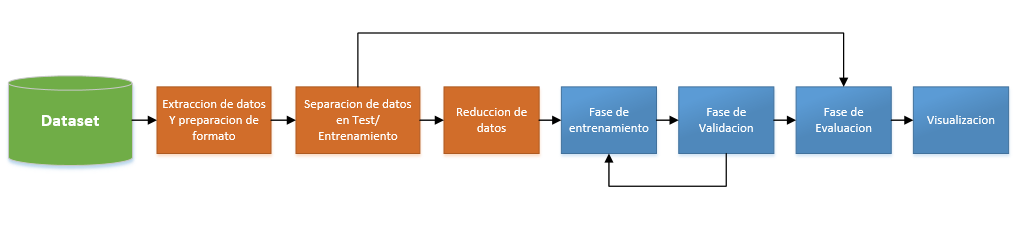


Ilustración 2.10 Esquema del proyecto. En naranja los temas que forman parte del capitulo 4 y en Azul aquellos que forman parte del capitulo 5.

En la figura 2.10 se tiene un esquema general de cuáles son las etapas del proyecto. Sin embargo, las figuras 4.3, 5.1 y 5.3 se tratan de esquemas que detallan cada una de las etapas que corresponden los procesos de reducción de datos, entrenamiento/validación y de evaluación, respectivamente.

# 

Marco Teórico

## 3.1 Imágenes de resonancia magnética

La imagen de resonancia magnética es una técnica usada en radiología para representar las imágenes de la anatomía y los procesos fisiológicos del cuerpo. Se utilizan fuertes campos magnéticos, gradientes de campo magnético y ondas de radio para generar las imágenes de los órganos del cuerpo. A diferencia de la tomografía axial computarizada no se utilizan rayos X que ejerzan algún tipo de radiación.

Con respecto al uso de la tomografía axial computarizada (TC), las MRI son más utilizadas en el campo de la investigación debido a que la primera resulta más sensible a la detección de la estructura ósea, sin embargo, no aporta el mismo contrate entre tejidos que aportan las MRI que pueden logran contrastar la materia gris, la materia blanca y el líquido cefalorraquídeo. Adicionalmente, como el intervalo entre una MRI y otra MRI es de milisegundos, es posible mostrar como el cerebro reacciona ante diferentes estímulos. En neuro-imagenologia las MRI son utilizadas para detectar demencia, enfermedades cerebrovasculares, síndrome de Alzheimer, epilepsias y, por supuesto, cáncer encefálico.

### 3.1.1 Tipos de MRI

#### 3.1.1.1 T1 y T2

Al aplicar campos electromagnéticos, un resonador magnético puede excitar los spines de las partículas de un tejido en particular si los campos electromagnéticos coinciden en el dominio de frecuencia con el valor de resonancia para un tejido. Los espines son excitados y los núcleos adquieren una carga magnética mientas que la frecuencia es emitida, esto afecta la posición original de los espines durante cierto tiempo.

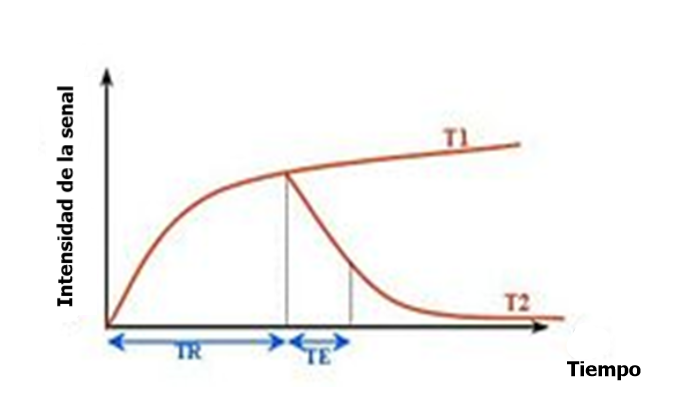


Ilustración 3.1 Diferentes tipos de relajación de un tejido especifico

El tiempo de relajación T1 es conocido como el tiempo que toma al spin regresar a su posición en el eje longitudal. Mientras que el tiempo de relajación T2 es el tiempo que toma al spin regresar a su posición original en el eje transversal. Estos valores no tienen ninguna proporción entre ellos y el valor de cada uno de ellos varía dependiendo del tipo de tejido. Por ejemplo, para la grasa el valor de T1 es alto en comparación a su valor de T2, sin embargo, en el caso de líquidos como los Edemas el tiempo T2 son mayores que su valor T1. [22]

#### 3.1.1.2 FLAIR

El acrónimo FLAIR viene del inglés, Fluid attenuated inversión Recovery. Es una secuencia basa en la técnica de inversión-recuperación [31]. La señal proveniente del líquido cefalorraquídeo es suprimida. En efecto, la señal emitida por un fluido o tejido en específico puede ser suprimida al escoger el tiempo de inversión específico del tejido o fluido de tal manera que sea filtrado de los datos originales. En general, se conoce que para el líquido cefalorraquídeo el tiempo de inversión es:

(3.1)

Es decir, el tiempo de inversión es aproximadamente un 70% del tiempo de relajación T1.

#### 3.1.1.3 Uso de agentes de contraste

Aunque bien las MRI permiten una manipulación de contraste a través del procesamiento de la imagen mucho más extensa que otros medios de radiología como la TC. Dado que es posible modificar el contenido químico de los tejidos para alterar sus resultados en MRI, para algunos problemas resulta necesario el uso de agentes de contrastes externos que son inyectados al paciente por vía intravenosa antes de realizar la imagen.

## 3.2 Cáncer encefálico

Los tumores, en general, se encuentran divididos en 4 grados de acuerdo a la forma de las células y el tejido de este sometido a un microscopio. Mientras más alto el grado el tejido tumoral tiene mayor similitud, a nivel celular, al tejido sano del cerebro. El grado es una indicación de la posible rapidez de crecimiento del tumor. [28, 29] Normalmente, los grados de los tumores son determinados después de realizar una biopsia, que son estudiadas por un patólogo. En el caso de tumores cerebrales estos se diferencian según el sistema de clasificación de la OMS, antes del año 2016:

* **Grado I (astrocitoma pilocitico):** Crecimiento lento, escasa tendencia a infiltrar en tejidos circundantes. El más frecuentes en niños y adolescentes.
* **Grado II (astrocitoma difuso):** Cierta tendencia a infiltrar en tejido cerebral circundante. Más frecuentes en adultos jóvenes.
* **Grado III (astrocitoma anaplasico/maligno):** Crecen con bastante rapidez con tendencia a infiltrarse en tejido cebreal circundante. Normalmente deriva de la malignizacion de un astrocitoma de bajo grado.
* **Grado IV (glioblastoma multiforme, GBM):** Es el caso más común y más maligno de tumores cerebrales. Representando el 67% de los casos. Afecta principalmente adultos entre los 45 y 70 años.

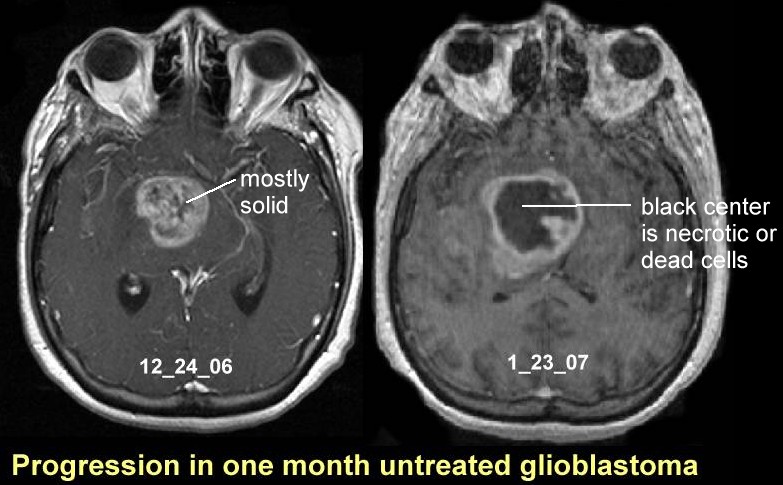


Ilustración 3.2 Progresión de un glioblastoma multiforme tras un mes sin tratamiento. [38]

Los gliomas de grado I y II son considerados como gliomas de bajo grado o, gliomas benignos. Los gliomas de grado III y IV son considerados como gliomas de alto grado o gliomas malignos. A partir del año 2016 la clasificación ha sido expandida considerando una clasificación más amplia y especifica de tipos y grados de gliomas. [30]

## 

## 3.3 Análisis de Discriminante de Fischer

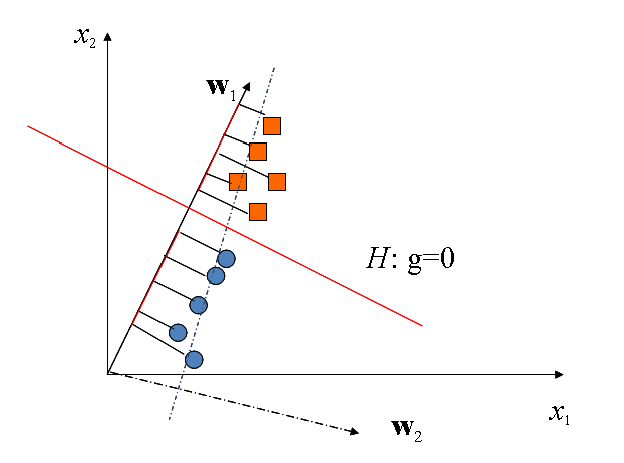


Ilustración 3.3 Proyección de un conjunto de datos con respecto a dos líneas diferentes. Se puede observar que la recta w1 muestra una mejor separación entre las dos clases de la imagen que la recta w2

El objetivo del análisis discriminante de Fischer es buscar la mejor dirección que pueda separar dos clases. [5] Para conseguir esta dirección, se supone que el set de datos D se divide en dos muestras D1 y D2 de acuerdo a las etiquetas de la clase. El objetivo será entonces de conseguir una línea y de forma:

Donde los puntos x que corresponden a D1 o a D2 estan bien separados por la dirección wd . El criterio normalmente utilizado para conseguir la mejor separación se define la siguiente función:

Donde se define la matriz dispersa de promedios entre clases:

Y la matriz de dispersa de covarianzas dentro de una clase:

De modo que el promedio de las muestras para una clase i es:

Y la covarianza para una clase i es:

La dirección óptima generalmente es calculada de la siguiente forma, en caso de que la proyección establecida para el análisis discriminatorio de Fischer se realice en un híper plano:

Sin embargo, para casos en que la proyección se realice en un espacio unidimensional se propone el radio de discriminante de Fisher:

Donde µ denota el promedio de cada clase. Mientras denota la variación estándar de cada clase.

## 3.4 Support Vector Machine

### 3.4.1 Support Vector Machine lineal

Se tiene un conjunto de muestras de muestras de forma donde x es la entrada estándar y los valores pueden ser 1 o -1 cada uno indiciando a la clase a la que podría pertenecer el punto , es decir, se realiza una clasificación binaria. Además, donde p representa el número de dimensiones y x es un vector. Se desea determinar el hiperplano de mayor cantidad de margen que logre separar los valores de los valores . Para esto, se determina el hiperplano tal que la distancia del más cercano de cada grupo sea maximizada. Cualquier hiperplano se puede describir como el conjuto de puntos que satisfaga:

Donde es un vector normal al hiper plano.

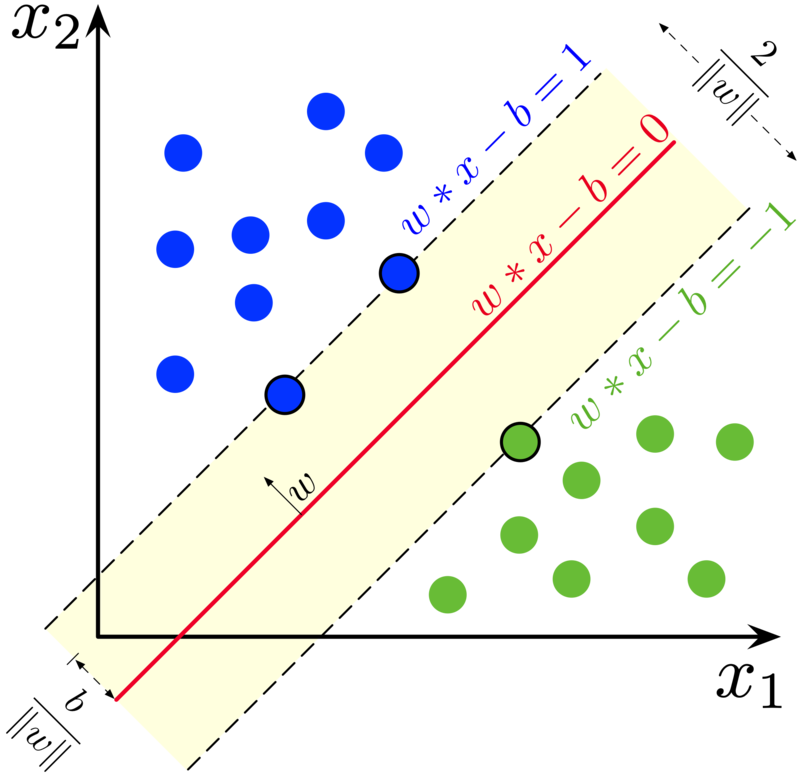


Ilustración 3.4 SVM en un dataset de 2D

Si los datos son linealmente separables, se pueden seleccionar dos hiperplanos paralelos que dividan los dos tipos de datos, tal que la distancia entre ambos sea la más larga posible. La región en medio de estos hiperplanos se considera “el margen” y el “hiperplano de máximo margen” es aquel que se encuentra en medio de los primeros hiperplanos mencionados. En un data set normalizado, estos hiperplanos se pueden definir desde las ecuaciones.

Los puntos a partir de este hiperplano son aquellos que cumplen esta ecuación pertenecen al grupo . Para el otro grupo de datos se tiene:

Pertenecientes al grupo .

La distancia entre los dos hiperplanos es por lo tanto, el ideal será de reducir . La distancia es calculada usando la ecuación de un punto a un hiperplano:

Donde y

Esto se reduce en:

Y el objetivo es el de reducir para cumplir

Finalmente, estos hiperplanos se determina utilizando los puntos más cercanos al hiperplano a determinar. Estos son llamados vectores de soporte.

### 3.4.2 Función de perdida para casos no lineales

Se introduce la función de *hinge loss*  (perdida de bisagra, en ingles):

Hay que remarcar que y es el output.

Esta función debería ser 0 para los casos lineales. Donde todos los puedan permanecer en el lado correcto del margen. Para los en el lado incorrecto del margen el valor es proporcional a la distancia entre el punto y el margen. Entonces, se propone optimizar la siguiente función:

+

Donde el parámetro es la perdida el incrementar el tamaño del margen. El valor de afecta directamente el rendimiento del SVM, por lo tanto es considerado como un hiperparametro. La aplicación de la función de perdida en los casos de SVM aporta a hacer la clasificación más robusta a outliers.

Se considera una muestra Z divisible entre X y Y. Siendo X un input estándar y Y el output escalar. Se considera la función perdida que mide el costo de predecir cuando el resultado es y se escoge una familia F de funciones que son parametrizadas por un vector . Se busca la función que minimice la perdida promedio en las muestras. Sin embargo, se plantea conseguir un promedio sobre la distribución desconocida de las muestras . Se definen las siguientes funciones

El riesgo empírico mide los resultados del entrenamiento del algoritmo mientras que el riesgo esperado mide los resultados en muestras futuras que resultan desconocidas. La teoría de aprendizaje estático justifica minimizar el riesgo empírico en vez de minimizar el riesgo esperado cuando se tiene una familia de funciones F estricta.

## 3.5 Gradient descent

Una de las formas de reducir el riesgo empírico es el gradient descent. Cada iteración actualiza los valores de w basándose en el gradiente de .

Donde es una ganancia elegida. Si se asume una regularidad en las muestras y el se halla lo suficientemente cerca de la optimización este algoritmo puede alcanzar una convergencia lineal

De esta forma se necesita una función diferenciable, para poder encontrar un valor mínimo de dicha función en todo su dominio. Se ejecuta un paso proporcional al sentido opuesto de la gradiente, por lo tanto, de ahí se deriva su nombre de gradiente descendente.

## 3.6 Stochastic gradient descent

El modelo estocástico del gradient descent responde al problema de elección de una muestra lo suficientemente cercana al estado óptimo y simplifica el modelo clásico al utilizar una muestra inicial aleatoria como base. Se propone la siguiente optimización de función de perdida para un clasificador:

Finalmente el proceso de esta función depende de las muestras aleatoriamente elegidas en cada iteración. Se espera normalmente que este proceso funcione a pesar del ruido que el procedimiento puede generar.

Gracias a que el proceso no tiene que almacenar las muestras que fueron visitadas en las iteraciones anteriores, puede procesar muestras con mayor velocidad. En esta situación, el proceso reduce el riesgo esperado ya que las muestras han sido tomadas de la verdad fundamental del dataset.

Sin embargo, la función de perdida de bisagra no es diferenciable. Sin embargo, es convexa y es posible generalizar las gradientes para dicha función, utilizando las subgradientes, es decir, la tangente de cualquier hiperplano que satisfaga la propiedad de estar bajo el punto . Formalmente, un vector es un subgradiente de para un punto si se satisface:

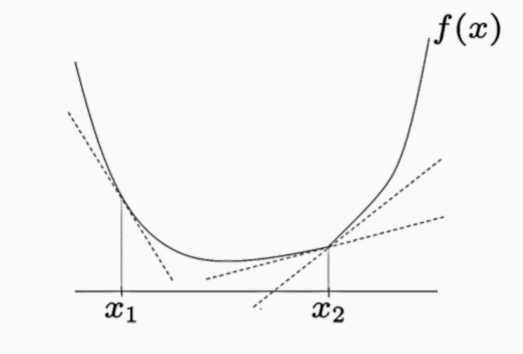


Ilustración 3.5 Demostración de un subgradiente. El punto X1 tiene una sola gradiente, sin embargo, el punto X2 tiene diferentes subgradientes.

La estrategia general en dado caso para la ecuación XXX es de resolver el valor máximo y calcular el gradiente para cada caso, es decir:

De esta manera se aplica el algoritmo de stochastic gradient descent, aplicado a un Support vector machine del siguiente modo.

1. Se inicializa el valor
2. Para cada iteraciones 1…T se toma para cada muestra

Si

Sino:

1. Devolver

## 3.7 Métodos de Kernel

Los métodos de Kernel permiten evitar los restricciones de los metodos de clasificacion lineal como lo son los SVM al ser utilizados en conjutos de datos que no son linealmente separables. Son utilizados para algoritmos donde se proponen dos vectores de entrada que dependen de un producto interno. Es decir, son funciones que toman los vectores de entrada y regresan el producto interno de los vectores en otro espacio dimensional.

Para cada y en el espacio de input Z, se definen funciones de forma que pueden representarse como el producto interno de otro espacio . Estas funciones se definen como . La computacion de los kernels se facilita al escribirla en forma de mapa de dimensionalidades lo cual satisface:

Le restricion principal es que sea un producto interno. Por un lado, no se necesita una representacion explicita de la funcion . Esto permite transformar un espacio dimensional a otro donde se pueda obtener una clasificacion lineal y este metodo es conocido como “el truco del kernel” [19]

Los diferentes tipos de kernel aplicados pueden variar la complejidad en tiempo y memoria del SVM o SGD de forma importante. Entre los mas destacados se encuentran el lineal, el polinomico y el RBF. El kernel lineal sigue representando una separacion lineal entre las dos clases, siento este el mas sencillo ya que no representa un verdadero mapeo de datos de un espacio vectorial a otro.

El Kernel polinomico, como su nombre bien lo indica mapea los datos a otro espacio vectorial de mayor dimensionalidad basandose en una funcion polinomica. De este modo es posible realizar clasificaciones mas complejas con fronteras no lineales. Es un kernel propenso al ruido y sensible al sobreentrenamiento. El uso de estos kernels implica una myor complejidad en memoria y tiempo.

El kernel RBF puede traza una frontera discontinua en el espacio de caracteristicas para mejorar la precision de la clasidicacion. Este kernel se define a partir de la funcion exponencial y puede realizar fronteras los suficientemente compleas, a pesar de la gran cantidad de complejidad de memoria y procesamiento que se requiere.

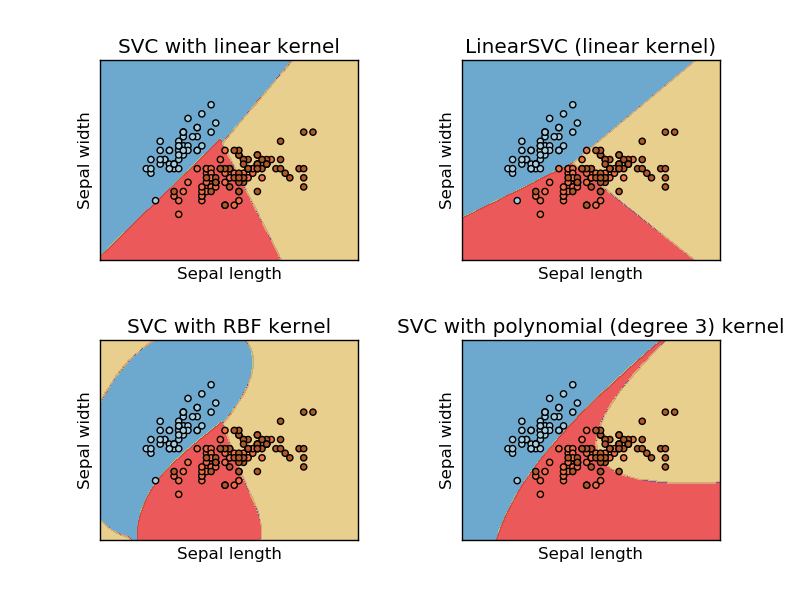


Ilustración 3.6 Algunos de los Kernels más utilizados [39]

## 3.8 K-Means

K-means es un algoritmo de agrupación no supervisado. El objetivo es de conseguir K grupos de datos a partir de una muestra determinada; el algoritmo funciona de forma iterativa, estimando K centroides inicialmente (estos pueden inicializarse de forma aleatoria). Se realiza el algoritmo de Esperanza – Maximización que es utilidado en una variedad de conceptos en la ciencia de datos. El algoritmo procede con los siguientes pasos:

1. Inicializar los clusters
2. Repetir los siguientes pasos hasta converger:
   1. Paso E (Esperanza): asignar los puntos al centro de cluster mas cercano.
   2. Paso M (Maximización): asignar el promedio de las posiciones de los puntos seleccionados en E como el nuevo centro de cada clúster.

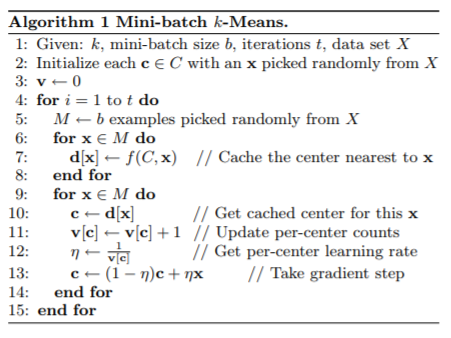


Ilustración 3.7 Evolución de K-means

La mayor desventaja de K-Means consiste en que para cada una de sus iteraciones es necesario calcular y almacenar la distancia de cada muestra con respecto a cada centro. Por lo tanto, el algoritmo se vuelve más complejo mientras mayor sea la cantidad de muestras y de clústers a asignar.

## 3.9 Minibatch K-means

Minibatch K-means es un algoritmo introducido en el 2015 por Chavan et al. [32] para poder superar los costos computacionales de K-means en grandes sets de datos y mejorar la complejidad de tiempo. El objetivo sigue siendo el de conseguir K grupos denominados clusters. Agregando el parámetro b que corresponde al tamaño de los lotes (mini-batches) donde para la etapa de Esperanza-Maximización se computarizaran la cantidad de b muestras aleatorias en el paso E, en vez de calcular las distancias de todos los puntos.



El algoritmo de mini-batch K-means aporta resultados similares a aquellos de los de la versión clásica de lotes completos de K-means. Adicionalmente, se sirve de la proximacion estocástica del algoritmo clásico. Su complejidad en tiempo es de O(kbmt), mientras que la versión original tiene una complejidad de O(tknm). Donde, t es la cantidad de iteraciones; k, la cantidad de clusters; n, el número de puntos; b, el tamaño de los lotes; y m la dimensionalidad de los vectores.

## 3.10 Análisis de componentes principales

El análisis de componentes principales [17] es de utilidad cuando se presentan un conjunto de muestras cada una fuertemente relacionada a variables diferentes a partir de m componentes tal que de tal manera de eliminar la redundancia al utilizar diferentes variables.

En la inteligencia artificial es una técnica de bastante utilidad ya que permite reducir los costos computacionales de procesar un conjunto de datos con muchas dimensiones.

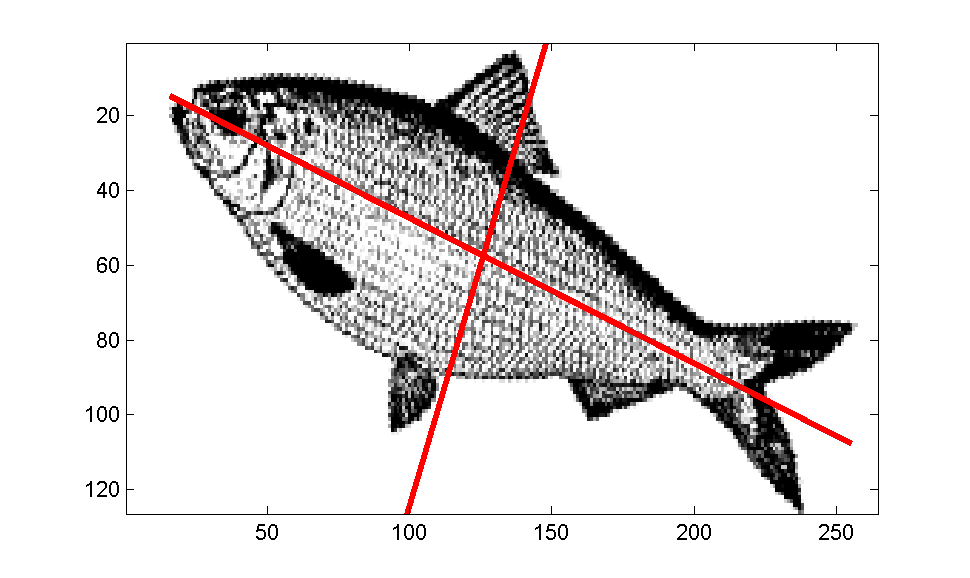


Ilustración 3.8 Los dos ejes de componentes principales en la foto de un pescado

El algoritmo detrás del análisis de componentes principales ejecuta los siguientes pasos:

* Tomar el set de datos original e ignorar las etiquetas del conjunto de datos. Donde d corresponde al número original de dimensiones.
* Hallar el promedio de cada dimensión del dataset.
* Calcular la matriz de covarianzas de toda la data set.
* Hallar los autovectores de los autovalores correspondientes.
* Ordenar los autovectores de forma decreciente y escoger K autovectores de tal manera de formar la matriz W de d x k dimensiones.
* Usar la matriz W de autovectores para transformar las muestras en un nuevo subespacio.

## 3.11 Sobreajuste

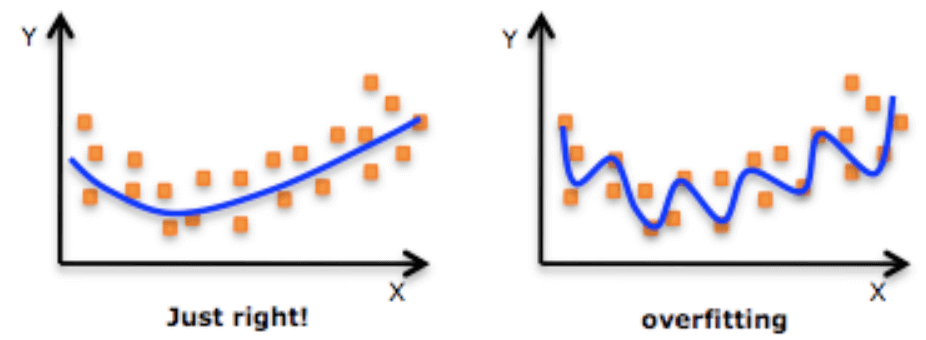


Ilustración 3.9. Ejemplo de un modelo de aprendizaje que representa sobreajuste

El sobreajuste u *overfitting*, es el efecto de sobreentrenar un algoritmo de aprendizaje supervisado, [9] si el modelo se ajusta de una forma muy cercana o exacta a un conjunto de datos particular y, por lo tanto, tiende a adaptarse a la variación residual o ruido que el conjunto de datos puede tener. Por lo tanto, un algoritmo de aprendizaje con sobreajuste puede fallar en generalizar un conjunto de datos para muestras futuras o desconocidas. Esto trae como consecuencia que al comparar los resultados de un algoritmo de aprendizaje con respecto a su mismo conjunto de entrenamiento se obtengan buenos resultados, sin embargo, al intentar predecir con respecto a un conjunto de datos diferente se obtienen peores resultados.

## 3.12 Validación cruzada

La validación cruzada o *cross-validation* en inglés, es una técnica para estimar el desempeño de modelos de aprendizaje con respecto a un conjunto de datos independiente, es decir, que no ha sido utilizado para entrenar el algoritmo con anterioridad y evitar el sobreajuste del modelo a un conjunto de datos de entrenamiento en particular. [18] También es utilizado para medir el desempeño de una serie de ajustes de hiperparametros, como por ejemplo, el número de iteraciones, una ganancia o el alfa de un SGD.

A diferencia de evaluar el rendimiento con una división de los datos entre los conjuntos de entrenamiento y prueba, se separa otro subconjunto de datos a partir del conjunto de entrenamiento, conocido conjunto de “validación”. Entonces, se entrena el algoritmo con el conjunto de entrenamiento, se realiza una evaluación con el set de “validación” para optimizar los hiperparametros y, al obtener resultados aceptables, se realiza una evaluación final con el conjunto de prueba. Sin embargo, dividir los datos en tres conjuntos reduce el número de muestras que pueden ser usadas para el modelo de aprendizaje y los resultados pueden depender de cómo se decidió dividir los conjuntos de entrenamiento y validación.

Por lo tanto, la validación cruzada toma el mismo set de datos de entrenamiento para usarlo como conjunto de validación, al aplicar el aprendizaje de forma iterativa. La forma más básica de realizarla, se conoce como *k-fold cross-validation*, el set de entrenamiento se divide en k subconjuntos y se aplica el siguiente procedimiento:

* El modelo se entrena usando k-1 subconjuntos de datos.
* El modelo resultante es validado y se toma como referencia un puntaje definido.
* El proceso se repite hasta lograr un total de k veces y se promedian los resultados como rendimiento global de los resultados para dicho puntaje definido.

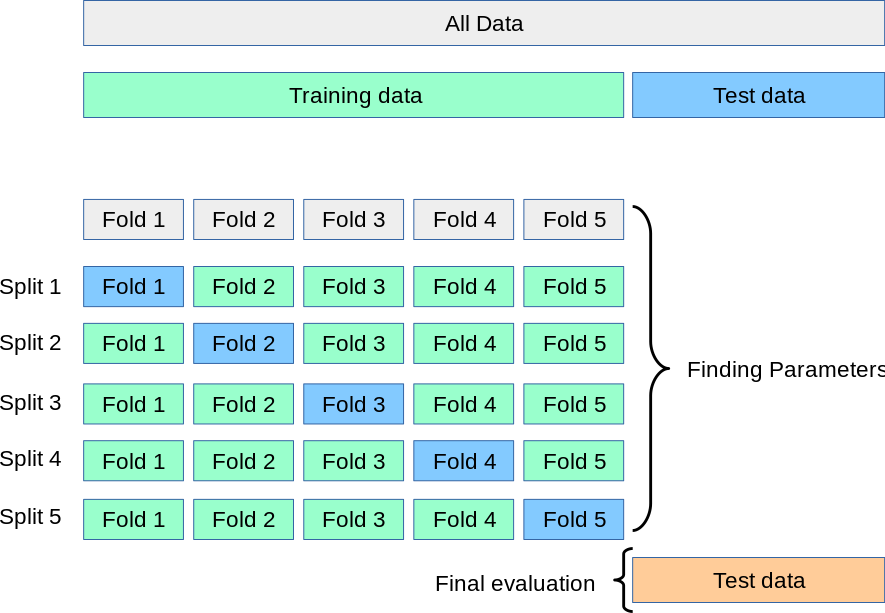


Ilustración 3.10. Ejemplo de una validacion cruzada.

## 3.13 Asimetría estadística

Cuando se desea analizar un conjunto de datos, la asimetría estadística y la curtosis aportan información sobre la distribución de los datos, sin necesidad de graficarlos. La asimetría estadística (o *Skewness* en inglés) representa, como su nombre lo indica, una medida cuantitativa para determinar la asimetría de la distribución de los datos con respecto a la media aritmética [40, 42]

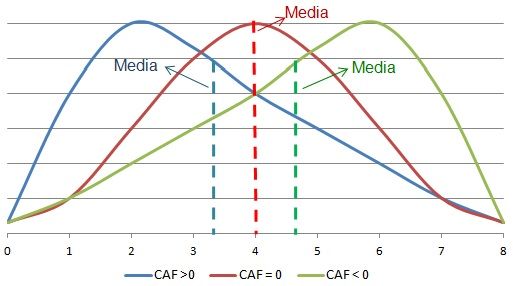


Ilustración 3.11 Diferentes valores de asimetria estadistica [42]

Existen tres tipos de distribución de datos, dependiendo de la asimetría estadística dependiendo del valor de la magnitud de la asimetría estadística:

* **Asimetría estadística nula:** En este caso, la mediana, la media y la moda coinciden en este punto y la distribución de los datos es similar a la distribución normal.
* **Asimetría estadística positiva:** La distribución de datos se alarga para valores superiores a la media.
* **Asimetría estadística negativa:** la distribución de datos se alarga para valores diferentes a la media.

Existen tres tipos de coeficientes para calcular la asimetría estadística: coeficiente de asimetría de Fisher, coeficiente de asimetría de Bowley y el coeficiente de asimetría de Pearson. En este estudio, se utilizó el coeficiente de Fisher el cual evalúa la proximidad de los datos a la media:

Donde representa la media y la desviación estándar.

## 3.14 Curtosis

La curtosis [41] se define mediante la siguiente formula:

Donde E representa el operador esperanza matemática y es el 4to momento centrado o con respecto a la media. Una distribución normal tiene una curtosis de 3, por lo tanto, en muchas ocasiones se utiliza para tener como el 0 de referencia a la curva de la distribución normal. En el caso de un conjunto de señales discretas permite entender la semejanza de la distribución de sus datos con la distribución normal.

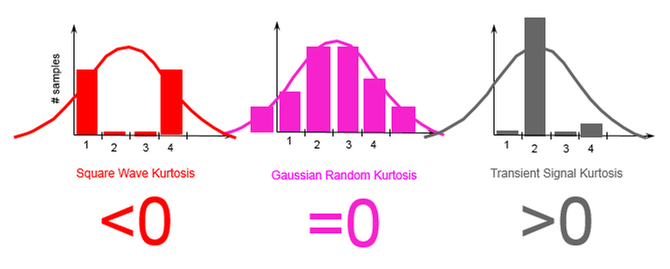


Ilustración 3.12 Variabilidad de una señal según su curtosis.

# 

**Procesamiento de datos**

En este capítulo se establecerá la metodología empleada para la preparación y reducción de datos. Todo esto después de analizar los métodos de reduccion y dar un análisis de por qué se decidió proceder con dicha metodología.

En la **sección 4.1** se presenta el dataset utilizado y cuáles son sus características más importantes. Adicionalmente, se establece como se definirán las etiquetas a predecir y qué tipos de tomografías serán utilizados. Mientras que, en la **sección 4.2,** se ofrece un análisis pre-exploratorio de dichos datos con el fin de entenderlos antes de aplicar cualquier algoritmo, en dicha sección, se exponen las problemáticas más importantes de los datos y un pre análisis de la factibilidad de aplicar métodos de clasificación de pixeles para obtener una clasificación aceptable.

En la **sección 4.3** se presenta el modelo con el que se pre procesaran los datos y en la **sección 4.4** se define como se han ordenado y seleccionado los datos que se utilizaran en las próximas secciones con el fin de reducir el desbalance de los datos.

De las **secciones 4.5** **a la 4.10** se detallan los pasos que se siguieron para reducir el volumen de los datos. Cabe destacar que este procedimiento de reducción está basado en el trabajo de Shen[37]. Sin embargo, se realizaron ciertas modificaciones las cuales también son justificadas en dichas secciones. Finalmente, en la **sección 4.11** se analizan los resultados del procedimiento de reducción, siendo ejecutado sin modificaciones y tras ser modificado.

## 4.1 Set de datos utilizado

El dataset MICCAI BRATS 2017 utilizado en la competición Brain Tumor Segmentación 2017 [7] [8]. Consiste en escaneos en 3D de encéfalos de 210 pacientes con Gliomas de alto grado y 75 pacientes con gliomas de bajo grado. Cada uno en 4 modalidades diferentes (T1, T1 con agentes de contraste, T2, FLAIR).

Los datos distribuidos ya han sido pre procesados al alinear las diferentes modalidades a la misma plantilla anatómica, interpolados a la misma resolución de 1 mm3 y el encéfalo ha sido separado de su cráneo para este estudio. Este último proceso se conoce como *skull-stripping*.

Adicionalmente, los voxeles fueron ya segmentados de forma manual por neuroradiologos profesionales, para dar un resultado de “verdad fundamental”. Donde, se aplican cuatro etiquetas:

* **Tumor realzado: (etiqueta 4.0)** Tejidos que asumen agentes de contrastes como el Gadolinio.
* **Edema peritumoral (etiqueta 2.0)** expresión de alteración de la barrera hematoencefálica.
* **Tumor necrótico y tejido tumoral no realzado (etiqueta 1.0)** tejidos que no reaccionan a partir de los diferentes agentes de contraste; debido a que existe un bloqueo en el flujo sanguíneo del tejido.
* **Tejido sano o nulo (etiqueta 0):** En esta categoría se incluye los voxeles que pertenecen a tejidos cerebrales sanos o que no pertenecen en lo absoluto al cerebro.

Es muy importante resaltar que durante el resto del trabajo se juntaron las etiquetas de la verdad fundamental de los edemas, el tumor realzado y el tumor necrótico en una sola. Ya que dividir el tumor segmentado en sus diferentes partes esta fuera del perímetro inicialmente planteado para este proyecto.

Por lo tanto, las etiquetas a utilizar, a partir del **capitulo 4.3**, son las siguientes.

* **Clase 0:** Tejido sano o nulo.
* **Clase 1:** Tejido anormal del cerebro.

Adicionalmente, se dejaron fuera del perímetro el estudio de los tumores de bajo grado incluidos en el dataset. Ya que estos casos son poco comunes en pacientes adultos, tal y como se explicó en el marco teórico.

## 4.2 Análisis pre-exploratorio de los datos

Comprender los datos con los que se trabaja es un paso esencial antes de ejecutar algoritmos de aprendizaje automático. Por lo tanto, se procede a estudiar todos los volúmenes de los análisis de los diferentes pacientes con tumores de alto grado.

Se realiza un análisis cuantitativo de los datos en la **sección 4.2.1** ya que, debido a su gran volumen y variabilidad, no es posible analizarlos individualmente en forma de histogramas. Por lo tanto, se propone analizar valores estadísticos para entender la evolución de las diferentes clases según cada modalidad de MRI. En la **sección 4.2.2**, se propone un diagrama de dispersión en 3D como una alternativa a los histogramas para poder visualizar los datos y dar a entender de forma intuitiva como estos pueden estar dispersos en un paciente aleatorio.

La **sección 4.2.3** sirve para exponer uno de los principales problemas del dataset: el volumen de los datos y, finalmente, la **sección 4.2.4** detalla otra problemática del dataset la cual es el gran desbalance que existe entre la cantidad de muestras pertenecientes al tejido sano con respecto al tejido tumoral.

### 4.2.1 Análisis cuantitativo de los datos

Para hacer un análisis cuantitativo de los datos se procedió a determinar individualmente para los volúmenes completos en 3D de los 210 pacientes con tumores de alto grado y para cada uno de ellos se halló:

* Promedio de valores en los voxeles
* Varianza
* Mediana
* Asimetría estadística
* Curtosis

Estos análisis se hicieron por cada etiqueta del dataset original y modalidad de MRI de forma separada; es decir, se analizó el tejido sano, el tumor necrótico, el tumor no realzado y el edema, para cada modalidad. Finalmente, se analizó también el conjunto de clases pertenecientes al tejido tumoral, como se determinó que se trabajaría en el capítulo anterior.

Los análisis estadísticos de los 210 pacientes se pueden conseguir en el **apéndice B** de este trabajo. La tabla 4.1 representa el promedio de cada uno de los resultados obtenidos de los resultados de cada paciente, con el fin de establecer como serían los datos de un paciente promedio del conjunto de datos.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Modalidad | Clase | Promedio | Varianza | Mediana | Asimetria Estadistica | Curtosis |
| T1 | Tejido Sano | 425.30 | 20623.93 | 440.14 | -0.71 | 1.65 |
| Tejido Tumoral | 423.19 | 7963.90 | 420.95 | 0.10 | 1.09 |
| Tumor necrotico | 385.05 | 5876.18 | 382.43 | 0.16 | 1.11 |
| Edema | 435.93 | 6029.08 | 435.28 | -0.05 | 0.70 |
| Tumor R | 423.40 | 5557.95 | 420.54 | 0.18 | 1.49 |
| T2 | Tejido Sano | 413.39 | 38956.14 | 373.54 | 1.49 | 4.05 |
| Tejido Tumoral | 591.91 | 28598.13 | 578.87 | 0.47 | 0.57 |
| Tumor necrotico | 709.94 | 27216.72 | 713.68 | -0.14 | 0.59 |
| Edema | 551.65 | 16216.07 | 546.06 | 0.34 | 0.25 |
| Tumor R | 598.42 | 18649.27 | 592.07 | 0.21 | 1.01 |
| T1C | Tejido Sano | 451.42 | 25414.10 | 459.25 | 0.83 | 9.96 |
| Tejido Tumoral | 500.08 | 44973.91 | 463.02 | 1.47 | 3.51 |
| Tumor necrotico | 425.82 | 13421.52 | 418.64 | 0.42 | 1.13 |
| Edema | 440.34 | 10826.69 | 437.47 | 0.47 | 2.74 |
| Tumor R | 721.29 | 34810.30 | 704.92 | 0.62 | 0.66 |
| FLAIR | Tejido Sano | 240.71 | 6064.43 | 250.62 | -0.53 | 2.85 |
| Tejido Tumoral | 424.24 | 9970.03 | 418.28 | 0.21 | 0.82 |
| Tumor necrotico | 435.57 | 8839.39 | 436.52 | -0.11 | 1.05 |
| Edema | 413.28 | 5379.05 | 411.61 | 0.13 | -0.01 |
| Tumor R | 433.45 | 9560.34 | 430.34 | -0.02 | 1.16 |

Tabla 4.1 Analisis estadistico de las muestras

A partir de estos datos se pueden establecer las siguientes observaciones:

Para **la modalidad T1:** En el caso del tejido sano con respecto al tejido tumoral, se puede observar que están ciertamente superpuestos, sin embargo, sus asimetrías estadísticas difieren completamente uno del otro. Esto significa que, para la mitad del área inferior hay mayor densidad de tejido tumoral, mientras que en la otra mitad se haya una mayor densidad de tejido sano. El tumor necrótico es el que más distancia tiene del resto, considerando que tiene una media de 385.05 y una mediana de 382.43.

Para **la modalidad T1 con agente de contraste:** Se denota cómo se reduce la diferencia entre los tejidos tumorales y no tumorales, donde, por ejemplo, el tejido sano, el tumor necrótico y el edema pueden tener cierta similitud en sus intensidades de pixeles. No obstante, el tumor realzado tiene características que difieren bastante del resto de los tejidos, esto no es sorprendente debido a que los agentes de contraste son elegidos con el fin de resaltar la parte del tumor correspondiente al tumor realzado y así trazar los límites entre los edemas y el tumor completo.

Para **la modalidad T2** El tejido tumoral y sano representa una separación importante de medias y medianas. Al tener el tejido sano una mayor asimetría estadística que el tejido tumoral se estima que puede haber una distancia importante entre los dos conjuntos, ya que el tejido sano es más denso en valores de intensidad menores, mientras representa una alta curtosis. En este caso, el tumor necrotico es aquel que representa una mayor separación del resto de los tejidos.

Para **la modalidad FLAIR**: La asimetría estadística del tejido sano es de signo opuesto al conjunto de tejidos tumorales. El tumor necrótico, es aquel que difiere más del tejido sano de acuerdo a la diferencia de sus medias y sus curtosis. Sin embargo, el edema, el tejido tumoral necrótico y los tejidos tumorales realzados tienen poca diferenciación entre sí. Nótese en este caso como la aplicación del FLAIR hace que el tejido sano del cerebro se aleje más del tejido tumoral.

Finalmente, de lo mencionado, se puede concluir que el tejido sano del cerebro está fuertemente separado del resto en la modalidad FLAIR; el tumor realzado, en la modalidad T1C; y el tumor necrótico, en las modalidades T1 y T2. Sin embargo, este no es el caso para el edema tumoral en ninguna modalidad y que, por lo tanto, resultaría posible dividir el tejido tumoral del tumor necrótico y realzado al utilizar métodos de clasificación de pixeles en al utilizar las 4 modalidades, sin embargo, es posible que ocurran problemas al clasificar correctamente a qué clase (tejido sano o tumoral) pertenece un pixel que pertenece al edema.

### 4.2.2 Análisis cualitativo de los datos

Debido a que realizar histogramas para 210 pacientes diferentes y 4 modalidades diferentes no permite tener suficiente información sobre la bimodalidad entre el tejido cancerígeno y sano. Se experimentó realizar un diagrama de dispersión en 3D con un paciente aleatorio para verificar la forma y la distribución que las diferentes muestras deberían tener en un cerebro con gliomas. Se tomaron como ejes las modalidades FLAIR, T2 y T1C.

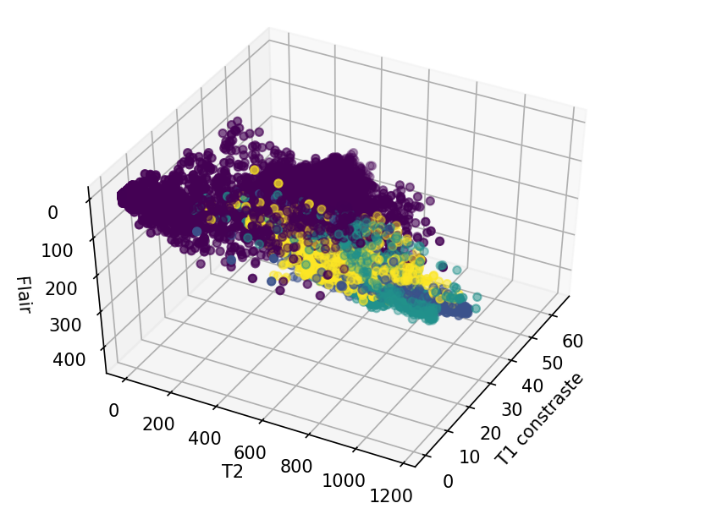


Ilustración 4.1 Diagrama de dispersión en 3D para un solo corte desde diferentes ángulos. Violeta: Tejido cerebral sano.

Se puede apreciar de las imágenes anteriores que es posible separar los tejidos sanos del cerebro de tejidos afectados por el cáncer en diferentes regiones basándose solamente en las características de los valores de la mayoría de los voxeles sin necesidad de considerar características espaciales. Sin embargo, se puede determinar que existe una superposición entre tejidos tumorales y sanos para algunos tejidos.

### 4.2.3 Dimensionalidad del data set

Los volúmenes de cada escaneo de resonancia magnética vienen definidos de acuerdo a las medidas: 240x240x160 dando un total de 8.920.000 voxeles por escaneo. Considerando que por cada paciente se ofrecen 5 modalidades diferentes (T1, T2, FLAIR, T1+C y RES). Se tendría un total de 44.640.000 datos por paciente. Finalmente, considerando la cantidad de pacientes de tumores de alto grado (210), la cantidad de datos total dentro del data set son 9.374.400.000.

Los archivos se encuentran almacenados en formato .nii.gz utilizado normalmente para las MRI. El peso total de los archivos de tomografías de alto grado es de un total de 1,6 GB. Sin embargo, el formato .nii.gz es la versión comprimida de los archivos. En total una resonancia magnética promedio por paciente tiene un peso de 16,5 MB, sin comprimir. Por lo tanto, siendo un total de 5 dimensionalidades por paciente y 210 pacientes, se estima un peso de 17,5 GB para el dataset original sin comprimir.

La infraestructura de hardware con la que se cuenta para aplicar algoritmos de inteligencia artificial no dará abasto para procesar esa cantidad de datos. Por lo tanto, el primer problema a solucionar es como reducir el data set con los datos que aporten la mayor utilidad a este estudio y eliminar los datos que aporten menor valor en nuestro dataset, debido a redundancias que se pueden generar al agregar datos de los mismos pacientes.

### 4.2.4 Proporcionalidad de tejido tumoral y tejido sano

Finalmente, para llegar a comprender cuantos voxeles, en promedio, hay en cada paciente para cada clase y que proporción establecen estos con respecto al tejido tumoral. Se extrajo la cantidad de voxeles por clases para cada uno de los 210 pacientes (Tabla también disponible en el **apéndice B**) y se promediaron. Finalmente, se dividió cada una de estas cantidades entre la cantidad de voxeles correspondiente a la unión entre las etiquetas de edema, tumor necrótico y tumor no realzado. Los resultados de dicho análisis se hallan en la tabla 4.2.

|  |  |  |
| --- | --- | --- |
| Clase | Cantidad Promedio | Cantidad/Tejido Tumoral |
| Tejido Sano | 1342953.46 | 13.95 |
| Tejido Tumoral | 96245.45 | 1.00 |
| Tumor necrotico | 14532.00 | 0.15 |
| Edema | 58689.05 | 0.61 |
| Tumor NR | 23024.40 | 0.24 |

Tabla 4.2 Proporcion entre tejidos tumorales y tejidos sanos en promedio por cada paciente.

En este caso, se puede observar que el tejido sano representa más de 13 veces el volumen que el tejido tumoral. Por otro lado, el tejido tumoral está compuesto en un 15% de tejido tumoral necrotico, un 61% de edema y 24% de tumor realzado. A partir de esto, se establece que es más efectivo entrenar el algoritmo de segmentación con los cortes en 2D que puedan aportar mayor tejido tumoral, el proceso se explica con detalle en la sección 4.4

## 4.3 Reducción propuesta

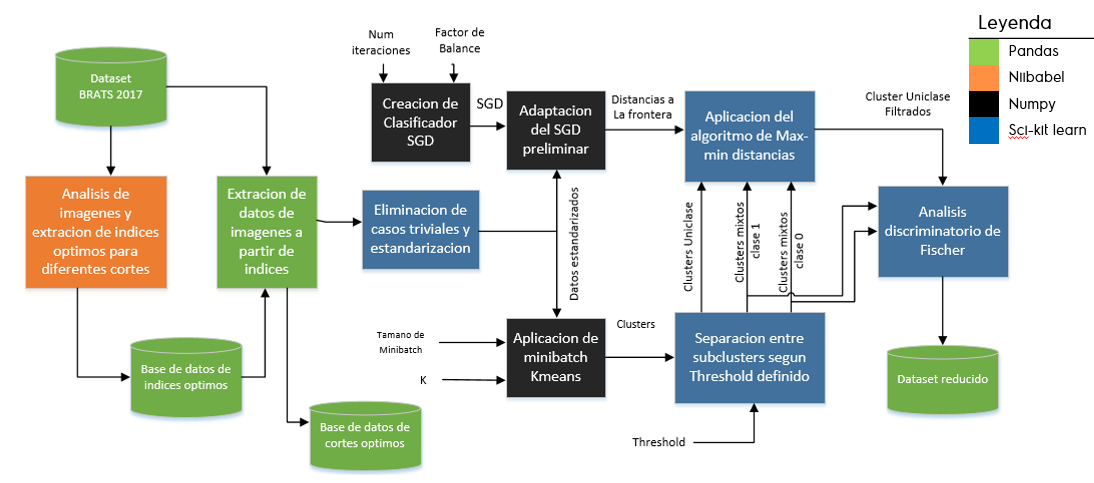


Ilustración 4.2 Diagrama de entrada y salida de los diferentes bloques funcionales de código.

Para reducir el dataset se propuso seguir el procedimiento propuesto por Shen et al (2016) [37] el cual está basado en el principio que estima que las muestras de interés para algoritmos de clasificación SVM son aquellos que se encuentran lo más cerca posible de la frontera. Se utilizan los siguientes pasos:

* **Eliminación de clústeres redundantes:** Implementar un algoritmo de agrupación (K-means, por ejemplo) para agrupar los resultados y eliminar los grupos que se encuentran lo suficientemente lejos del hiperplanos.
* **Eliminación de muestras redundantes dentro de los clústeres:** Implementar el análisis discriminante de Fischer para eliminar los puntos más cercanos al centro de los clústeres restantes del paso anterior y quedarse con aquellos puntos que representen la capa exterior del clúster.

Es importante resaltar, que se realizaron cambios importantes respecto a la propuesta inicial del autor. Las etapas de la reducción de datos aplicadas son explicadas, al mismo tiempo que aquellas que el autor propuso originalmente, con detalle en las próximas secciones.

## 4.4 Creación de una dataframe de índices de cortes óptimos y procesamiento de la base.

Para comenzar con una reducción inicial de datos y así poder procesar los datos que puedan aportar el mayor valor posible al esquema de aprendizaje, se decidió tomar las coordenadas óptimas para cada corte de un paciente en sus tres diferentes perspectivas (Coronal, Horizontal y Sagital), en un *dataframe* de pandas cuyo formato puede visualizarse en la tabla 4.3. Donde cada columna simboliza una coordenada específica de cada corte. Mientras que cada entrada simboliza un paciente particular.

|  |  |  |  |
| --- | --- | --- | --- |
| Nombre  Paciente | Coordenadas  Coronales | Coordenadas  Sagitales | Coordenadas  Horizontales |

Tabla 4.3 Encabezado del dataframe de coordenadas de cortes optimos.

Se considera una coordenada óptima para un plano de una MRI la posición del corte que en dicho plano donde se obtenga el máximo valor de voxeles etiquedos para la clase 1, 2 o 4. Este proceso es repetido para cada una de los tres planos correspondientes a cada modalidad y es posible realizar dicho proceso solamente con los datos de la verdad fundamental del paciente.



Ilustración 4.3 : Tres cortes principales de una tomografía

Una vez los datos son almacenados. Se generaron otras funciones diferentes para extraer los valores ordenados de cada pixel en cada una de las cuatro modalidades de MRI disponibles (T1, T1C, T2 y FLAIR) y los resultados de la verdad fundamental. Entonces, son ordenados en otro dataframe de Python como el de la tabla 4.4. Para este dataframe, la columna “Nombre” será opcional al momento de realizar predicciones una vez los algoritmos de inteligencia artificial son entrenados.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Nombre | T1 | T1C | T2 | FLAIR | RES |
| Paciente 1 |  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
| … | … | … | … | … | … |
|  |  |  |  |  |  |
| Paciente 210 |  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

Tabla 4.4 Dataframe de datos de voxeles para entrenar el algoritmo en amarillo voxeles del corte coronal; en azul, corte sagital y en verde, corte horizontal. Cada celda representa, en realidad, todas las entradas representadas por corte optimizado de paciente.

En la tabla mencionada, como los tamaños de los volúmenes de tomografía son ya conocidos y que cada uno corresponde a un tipo de corte particular. Para cada una de las celdas de color amarillo o azul representa, en realidad, un arreglo de 38.400 x 1 muestras y para las celdas de color verde se representa un arreglo de 57.600x1. Finalmente, por paciente y modalidad se extraen 134.000 muestras.

## 4.5 Estandarización de datos y eliminación de outliers.

Un procedimiento critico a realizar antes de aplicar algoritmos de optimización estocástica, como en el caso de SGD y minibatch K-means, es el de estandarizar las diferentes características del conjunto de datos. Esto es importante, ya que el rango de valores para cada dimensión no son necesariamente los mismos y, en ocasiones, pueden estar fuertemente desproporcionados entre sí. En consecuencia, puede haber ruido y desbalances en los algoritmos de optimización estocástica porque, al funcionar de forma iterativa basándose en las gradientes, es posible que haya una característica de la gradiente con un peso mucho más elevado que el otro, haciendo que sea posible alejarse innecesariamente del mínimo local.

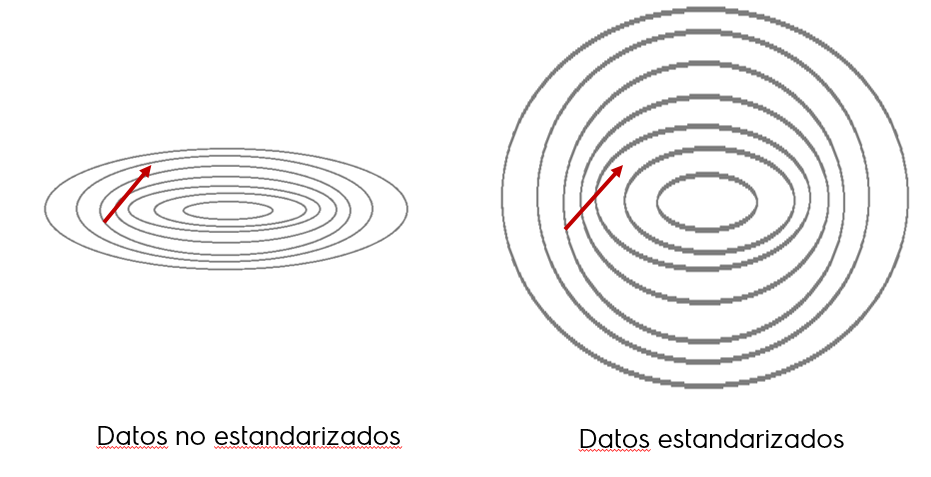


Ilustración 4.4 Se muestra una iteración de SGD para el mismo dominio sin estandarizar y tras ser estandarizado. Se puede apreciar cómo para el mismo punto y misma gradiente, resulta más óptimo estandarizar las dimensiones de los datos.

Para reducir este riesgo, se estandariza cada dimensión a un conjunto de datos con promedio de cero y una varianza unitaria. Una técnica de estandarización sencilla es reemplazar las variables por su puntaje z que se define para una dimensión:

Donde S es la variación estándar, µ el promedio y x es el valor de una muestra dentro de la dimensión especifica. Adicionalmente, pueden considerarse como outliers aquellas muestras que están fuera del rango [-3, 3] en cualquiera de las dimensiones debido a que están muy alejados de la distribución normal del conjunto, como explica Cousineau en su estudio [36].

No eliminar estas muestras puede interferir en los próximos algoritmos a utilizar, especialmente en el uso de minibatch k-means, donde elecciones de outliers dentro del conjunto de datos aleatorios para el minibatch aportan alteraciones en las distancias calculadas.

El enfoque de este trabajo es sobre lo que ocurre en la región de frontera del algoritmo de SGD, por lo tanto, se puede considerar eliminar los outliers para esta etapa. En pruebas empíricas, se observó que existen valores de outliers que pueden afectar el desempeño de la agrupación de clústeres y de la aplicación del algoritmo de *MAX-MIN distance*.

## 4.6 Aplicación de agrupación inicial

Para agrupar los datos, debido al gran volumen del dataset, el algoritmo clásico de K-Means resulta ser un proceso lo suficientemente lento [1, 32]. Considerando que se tendrían que calcular la misma cantidad de distancias euclidianas de cada muestra al centro del clúster, durante cada iteración. Es decir, se tendrían que almacenar y actualizar por iteración , variables donde es el número de dimensiones y el número de dimensiones del dataset.

Por lo tanto, se propone utilizar el algoritmo minibatch K-means, para reducir el número de muestras que se deben procesar por iteración a un número arbitrario. El aporte de Shen et al. (2016) [37] propone escoger un numero de K lo suficientemente grande para obtener clúster unitarios. Se propone K = 50 para esta implementación. En cuanto al tamaño de minibatch se plantea 50mil muestras para tener, en promedio, 1000 muestras siendo procesadas por clúster.

### Asignación de subconjuntos a los clúster.

El siguiente paso propuesto tras la aplicación de K-means, en el trabajo original, es el de clasificar los clústeres en los siguientes dos conjuntos:

* . **Clústeres unitarios**, los clústeres que tienen todas las muestras que han sido agrupadas pertenecen a la misma clase. Estos clústeres serán sometidos a una posible eliminación tras aplicar el algoritmo de *Max-Min distance* entre clúster.
* . **Clústeres mixtos,** El conjunto de clúster donde hay muestras de diferentes clases. Cada clúster es divido, al mismo tiempo, en subconjuntos pertenecientes a la cantidad de clases a implementar. En el caso particular de este estudio y como se definió al comienzo de este capítulo los dos sub-conjuntos serian:
  + el conjunto de valores de la clase 0 (tejido no tumoral) dentro del cluster MC correspondiente.
  + el conjunto de valores de la clase 1 (tejido tumoral) dentro del cluster MC correspondiente.

Sin embargo, para un dataset de aproximadamente 20 millones de muestras y tras el ruido que se obtiene al sobreponer los resultados de varias imágenes tumorales entre sí. Es imposible obtener un clúster que pueda ser clasificado directamente en el conjunto UC. Por lo tanto, se propone definir una nueva condición para que haga que los criterios de pertenencia al conjunto UC sean menos selectivos:

Para cada clúster se tienen muestras para las cuales están clasificadas en clases tal que donde , se calcula entonces la cantidad de muestras tal que y se dividen entre el conjunto total de muestras n tal que . De esta manera, se puede calcular la proporción de muestras clasificadas como no tumorales respecto al resto del clúster.

Así, se define un umbral y los clústeres cuyo p cumplen con la condición son asignados al conjunto MC, mientras que aquellos que si la cumplen serán asignados al conjunto MC. De este modo, es posible definir un clúster como unitario dependiendo un margen deseado. Como en este caso la proporción de muestras con clase 0 es aproximadamente 5 veces mayor que aquellas de clase 1. Se decide usar un valor para T entre 0.05 y 0.1.

### 4.8 Inicialización de SGD

La propuesta de reducción de datos de Shen et al (2016) expone la necesidad calcular la posición aproximada de los hiperplanos que separarían los dos conjuntos. Aunque la propuesta original propone calcular la distancia de cada punto con respecto al hiperplano a partir de la siguiente ecuación.

Donde es el clasificador de una muestra y la distancia ecludiea del W que representa el hiperplano. Se propone originalmente algún método que pueda estimar un los parámetros de un SVM como SMO (Sequential Minimal Optimization) que viene implementada en la librería LIBSVM.

Por lo tanto, se decidió inicializar preliminarmente un clasificador SGD ya que este es el mismo algoritmo que será implementado en el capítulo IV. Se decidió someter los puntos x a la función de clasificación del SGD (ya inicializado y entrenado) debido a que este valor cumple con la condición de ser directamente proporcional a la distancia de dichos puntos al hiperplano. Respecto a los hiperparametros del SGD a implementar, no se realiza una optimización estricta ya que este primer uso de SGD es solamente referencial para obtener una distancia inicial y no nos interesa establecer una verdadera clasificación por los momentos.

## 4.9 Eliminación de clústeres redundantes usando Max-Min cluster Distance

Para la reducción de clústeres redundantes, en el esquema propuesto, se usó el algoritmo de distancias Max-Min de clúster [37]. Donde se propone el siguiente seudocódigo equivalente para solamente 2 clases:

1. i = 0
2. **while** (i < 2) **do:**
3. **for** (cada cluster ) **do:**
4. **For** (cada punto j en ) **do:**
5. Calcular las distancias a partir de la función definida (en este caso la función de decisión del SGD)
6. **End for**
7. Asumir que el conjunto de distancias en el clúster es DDK; guardar la distancia más grande en el conjunto como la distancia del clúster
8. **End for**
9. Al haber calculado la distancia más grande de cada clúster, reservar la mayor distancia entre los clústeres de cada clase y asumir esta distancia como CMaxi
10. **For (**Cada clusters UCj en UC cuya clase es j) **do:**
11. **For** (cada punto m en UCj) **do:**
12. Calcular las distancias de cada punto hasta el hiper plano. Reservar la distancia más corta entre todas y llamarla X
13. **End for**
14. **if** (X > Cmaxi):
15. Eliminar el cluster
16. Else:
17. Reservar el cluster
18. End for
19. Devolver el conjunto de clusters que no han sido eliminados

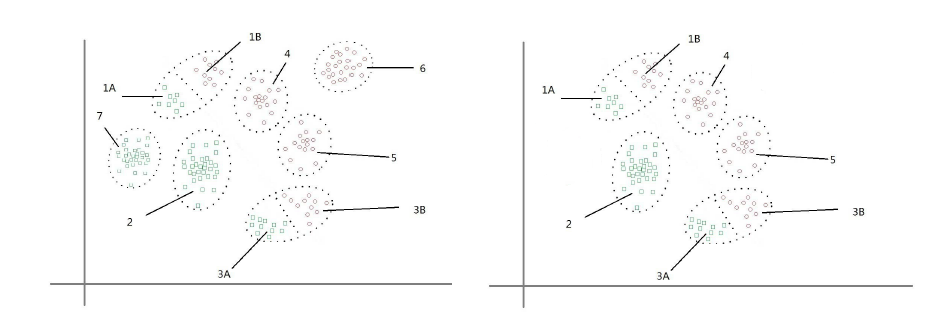


Ilustración 4.5 : Eliminación de clusters redundantes.[37]

En resumen, esta etapa del algoritmo consiste en eliminar por completo aquellos clústeres unitarios cuyos datos se encuentren más alejados de un hiperplano provisional definido tras la aplicación preliminar del SGD. Debido a que los datos que resultan de mayor interés son aquellos que se encuentran los más cerca posibles con la frontera con el hiperplano.

## 4.10 Reducción de datos usando FDA

En el procedimiento presentado se toma cada uno de los clústeres que fueron obtenidos como output de la eliminación anterior. La idea de esta etapa es la de reducir los datos más densos de cada clúster, es decir aquellos más próximos al centro. Se asumen los datos cercanos al centro de un clúster como agrupados, mientras que aquellos que están los suficientemente alejados del centro se definen como dispersos. Asumiendo que generalmente las fronteras de un algoritmo como SVM, se deberían encontrar a las afueras de clúster de una sola clase, se puede inferir que los datos agrupados de un clúster son redundantes.

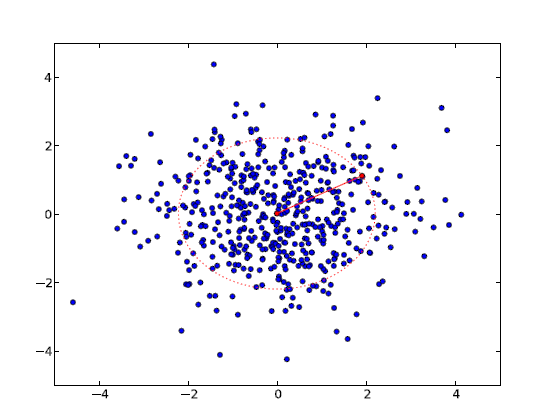


Ilustración 4.6 : Eliminación de datos redundantes en un solo clúster unitario. [37]

La lógica general para eliminar estos datos usando FDA, es la de calcular el mejor coeficiente de Fischer. Sin embargo, la recta que se utilizara para calcular los FDR, son en realidad, la recta entre cada muestra y el centro; mientras que las dos clases diferentes a separar son las muestras dentro del radio generado por dicha distancia, respecto a las muestras exteriores a ese radio. En la figura 4.7 se puede observar la visualización de este algoritmo.

Partiendo de una definición más formal, para un clúster unitario (Es decir, que los datos pertenecen a una sola clase) específico, se calcula el centroide del cluster, definido de la siguiente manera:

El siguiente paso es calcular las distancias de cada punto con respecto al centro del clúster , se puede tomar como medida de distancia la distancia euclidiana. Luego, se ordenan las muestras de menor a mayor según sus respectivas distancias al centro de su respectivo cluster. De forma que el conjunto de distancias ordenadas se denomine y la muestras correspondientes a dicha distancia se denotan por

Para un punto que corresponde a un valor . Se divide el set de muestras en dos, denotándolos como y y el conjunto de distancias en dos y donde los subconjuntos y representan las muestras y las distancias al interior del radio entre el centro del clúster y el punto , respectivamente; mientras que los subconjuntos y representan las muestras y distancias que se encuentran al exterior de dicho radio, respectivamente.

Una vez seleccionado un punto y establecido sus respectivos subconjuntos se desea calcular el FDR. El cual se halla a partir de la siguiente ecuación:

Donde y representan las segmentación los promedios de los valores en y , respectivamente. Mientras que y son las varianzas de y ; Se define entonces la barrera óptima de la siguiente forma:

Donde los puntos que pertenecen al del r óptimo son finalmente eliminados.

Adicionalmente, el autor propone un método de iteraciones rápidas para realizar la segmentación llamado FIFDR (*Fast iteration Fischer Discriminatory Ratio*) proponiendo el siguiente esquema de seudocódigo.

1. Calcular la variable de magnitud donde lo que n es la cantidad total de muestras para el clúster especifico y un número máximo de iteraciones
2. Calcular un y un que se inicializa como y .
3. **While( IL < IM ) do:**
4. Calcular el intervalo de pasos iterativos
5. **For ( do:**
6. Calcular el FDR del punto r; es decir, hallar , como ha sido explicado anteriormente.
7. **End for**
8. Seleccionar dos valores más altos hallados en el ciclo anterior. Suponiendo que es la muestra que devuelve el valor más alto y sea la muestra que devuelva el segundo más alto. Entonces se actualizan los valores , y
9. **End while**
10. Eliminar el conjunto y regresar como resultado los puntos .

Sin embargo, ha sido necesario adaptar el algoritmo propuesto al alterar el paso número 4 donde la asignación de valores para SL, no es la conveniente, por lo tanto se cambió la condición aplicada por donde la función representa el suelo de la parte entera, debido a que esta converge en los valores Bmax a partir de la primera iteración y converge a 1 en la última. En la fig 4.8. Se puede observar como se ha cambiado el algoritmo y como afecta su lógica.

Adicionalmente, ha sido necesario modificar la condición final donde se toman los valores más grandes de FDR por tomar los menores valores de FDR y realizar el corte a partir de dicho punto. Se darán más detalles sobre el comportamiento de esta función que hicieron necesario este cambio en los resultados de este capítulo.



Ilustración 4.7 : Modificación del algoritmo FIFDR propuesto originalmente por el autor. A la izquierda el esquema original, a la derecha la parte modificada. En naranja la etapa que ha sido modificada.

## 4.11 Resultados y análisis

La reducción de datos en las diferentes etapas del código es presentada en la siguiente tabla. Se analizaron la cantidad de datos restantes tras cada etapa y la proporción de tejido tumoral/no tumoral con el fin de conocer la evolución del balance de muestras. Se realizó la reducción de métodos tanto con el método propuesto por Shen et al (2016), como por el método modificado.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Clase | Dataset Inicial | Estandarización y eliminación de outliers | Eliminación de clústeres redundantes | Eliminación de datos redundantes en Clúster (FIFDR) |
| No tumoral | 8 599 476 | 8 180 534 | 6 147 241 | 2837983 |
| Tumoral | 1 527 938 | 1 385 578 | 1 318 190 | 714382 |
| Total | 10 127 414 | 9 566 112 | 7 465 431 | 3552365 |
| Proporción | 5,62:1 | 5,90:1 | 4,66:1 | 3,97:1 |

Tabla 4.5 : Reducción de datos a lo largo del proceso propuesto aplicando un Threshold de 0.1 en la eliminación de clústeres redundantes.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Clase | Dataset Inicial | Estandarización y eliminación de outliers | Eliminación de clústeres redundantes | Eliminación de datos redundantes en Clúster (FIFDR) |
| No tumoral | 8599476 | N/A | 8599476 | 6 389 410 |
| Tumoral | 1527938 | N/A | 1527938 | 1 152 065 |
| Total | 10 127 414 | N/A | 10 127 414 | 7 541 475 |
| Proporción | 5,62 :1 | N/A | 5,62 : 1 | 5,54:1 |

Tabla 4.6 Reducción de datos a lo largo del proceso propuesto por Shen aplicado sobre el mismo dataset.

Tras analizar la eliminación de outliers .Se destaca que inicialmente se redujo un 3,91% de los datos iniciales, debido a que estos son outliers. Como se mencionó anteriormente, esta eliminación de datos es necesaria ya que los outliers, pueden afectar seriamente el desempeño del *mini-batch K-means*, porque este es aplicado con métodos estocásticos. Al mismo tiempo, se asume que eliminar estos datos no representa una alteración al algoritmo de aprendizaje; debido a que estos valores se encuentran lo suficientemente alejados de la frontera entre las dos clases a analizar. Esta etapa no es propuesta en el esquema inicial de Shen (2016) [37].

Respecto al proceso de clusterizacion se puede observar que el algoritmo propuesto por Shen es muy estricto en cuanto a la clasificación de los clústeres en subconjuntos unitarios o mixtos. Se tiene que, en la tabla 4.7, todos los clústeres tienen al menos una muestra para cada clase. Por lo tanto, ningún clúster es considerado para una posible eliminación, en consecuencia, la eliminación es nula. Por otro lado, se resaltan los clústeres que se consideran dentro del rango si se toma un , como se recomendó en la **seccion 4.7**.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| 0 | 0.9414 | 0.7216 | 0.8110 | 0.9286 | 0.9699 | 0.7173 | 0.7828 | 0.8672 | 0.9886 | 0.2091 |
| 10 | 0.9296 | 0.1314 | 0. 9508 | 0.7557 | 0.9936 | 0.7653 | 0.6071 | 0.6506 | 0.5946 | 0.9051 |
| 20 | 0.3086 | 0.63175 | 0.6968 | 0.9764 | 0.9682 | 0.9747 | 0.9635 | 0.5407 | 0.7785 | 0.7284 |
| 30 | 0.9505 | 0.9873 | 0.9081 | 0.8884 | 0.8608 | 0.1548 | 0.9511 | 0.4795 | 0.9836 | 0.4885 |
| 40 | 0.9568 | 0.7676 | 0.9116 | 0.9603 | 0.8361 | 0.1755 | 0.6871 | 0.9234 | 0.9936 | 0.9959 |

Tabla 4.7Proporcion entre clases para cada cluster. Los clusters resaltados representan clusteres unitarios y seran candidatos para el algoritmo MAX-MIN Dist.

Tras la preselección de clústeres unitarios, se procedió a dividir los clústeres mixtos según las dos clases y a calcular cuales son las mayores y mínimas distancias, respecto al hiperplano del SGD que ha sido entrenado preliminarmente.

En la tabla 4.8, se destacan los clústeres que han sido considerados como unitarios y cuales han sido sus mínimas distancias con respecto al hiperplano. Entonces, la distancia máxima hallada en los clústeres mixtos ha sido de 1.661 (Clúster número 4). El proceso de eliminación de clústeres redundantes elimino 28.93 % de datos con respecto a la etapa anterior.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Cluster | Dist Min | Cluster | Dist Min | Cluster | Dist min |
| 0 | 1.394 | 23 | 1.929 | 38 | 3.057 |
| 3 | 1.372 | 24 | 1.509 | 40 | 2.461 |
| 4 | 1.661 | 25 | 2.000 | 42 | 1.718 |
| 8 | 0.980 | 26 | 1.292 | 43 | 1.576 |
| 10 | 1.745 | 30 | 1.934 | 47 | 2.208 |
| 12 | 1.570 | 31 | 1.501 | 48 | 2.082 |
| 14 | 1.943 | 32 | 3.551 | 49 | 2.278 |
| 19 | 1.489 | 36 | 1.415 |  |  |

Tabla 4.8 Clusteres redundantes eliminados tras la aplicación del algorimo MAX-Min distancias..

Respecto al análisis discriminatorio de Fischer, en la fig. 4.9, se analiza el comportamiento de los diferentes elementos que componen la función del FDR en un clúster aleatorio.

Nótese que para todos los valores de R se observa un valor de discriminante mayor que 1, es decir, todos los datos pueden ser linealmente separables. La cause es que, para este caso, todos los puntos son proyectados radialmente entre el centro del clúster y R, por lo tanto, las distancia del conjunto siempre son menores a aquellas del conjunto . Se puede observar el promedio para ambos conjuntos Xc y Xs a lo largo del dominio R.

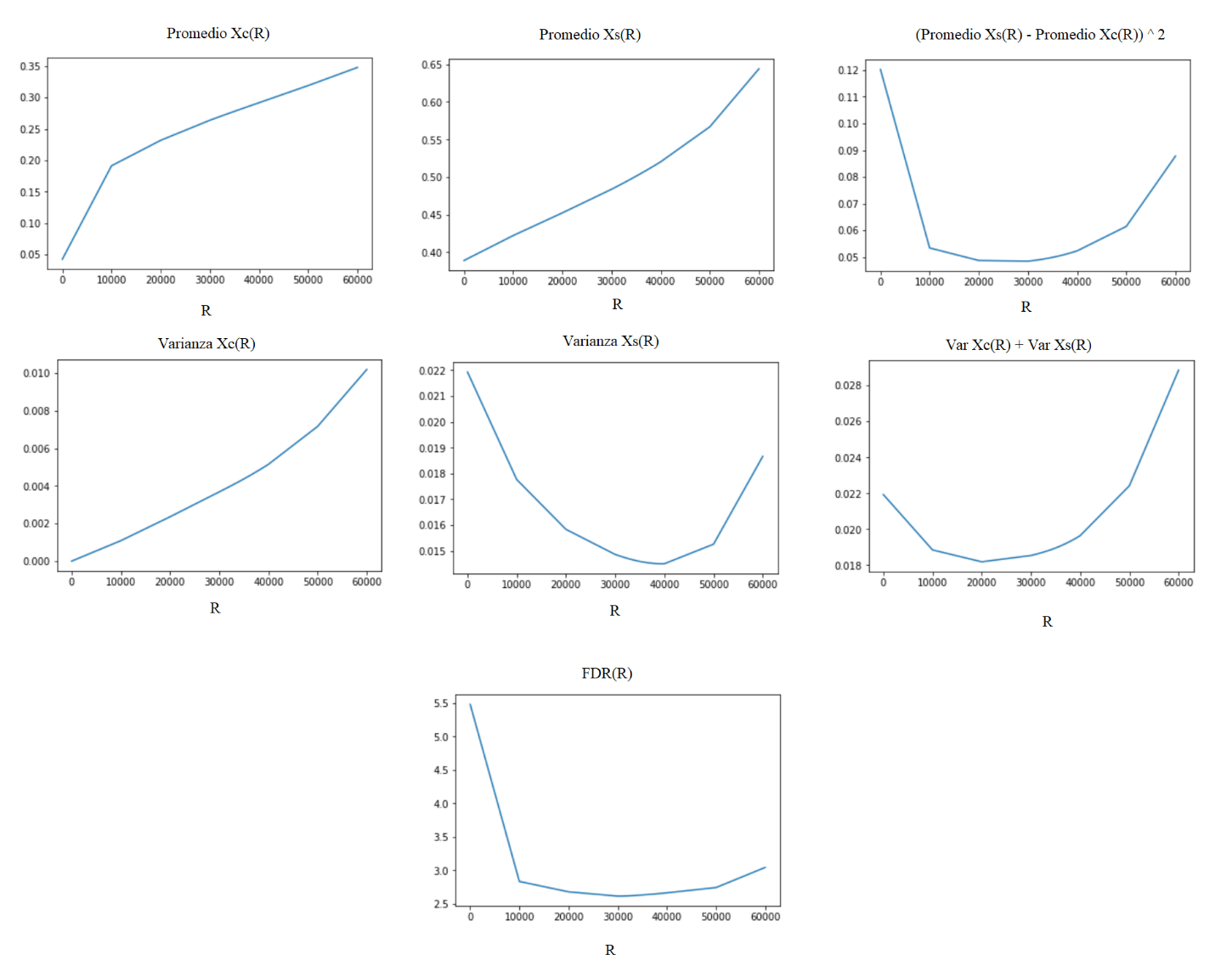


Ilustración 4.8 Comportamiento del FDR en un clúster aleatorio y sus diferentes parámetros.

Como se puede observar el mayor valor para el FDR a lo largo del dominio R, se puede conseguir en R = 1 o en R = n para la mayoría de los casos. Sin embargo, tomar alguno de estos dos valores implica no eliminar ningún dato del clúster o eliminar el clúster por completo. Por lo tanto, considerando que FDR > 1 para todo R y que, por lo tanto, ambos conjuntos resultan ser linealmente separables, se decidió tomar el valor de R que otorgué el menor de los FDR. De esta manera se logró una reducción de un 26.43% de los datos con respecto a la etapa anterior.

Aplicando la reducción recomendada, se logró reducir el tamaño del conjunto de datos basado en los cortes óptimos en hasta un 35,07% y el balance paso de ser 5,62:1 a 3,97:1. Mientras que el modelo de reducción de Shen [37] solo redujo el conjunto de datos a un 74,4% del valor original y la proporción a 5,54. Por lo tanto, la reducción recomendada resulto ser más eficiente, en termino de datos reducidos y balance de las clases.

Finalmente, considerando que el dataset tiene un conjunto de voxeles iniciales de 193.536.000 y que, después de todo el proceso de reducción de datos recomendado, se llega un total de 3.552.365 muestras. Finalmente, el algoritmo implementado ha seleccionado el 1,83% de las muestras más relevantes del conjunto original.

Se deja el desempeño de este nuevo conjunto de muestras como entrada del algoritmo de aprendizaje en el próximo capítulo, donde se detallara su funcionamiento.

# CAPITULO IV

**Entrenamiento del modelo de segmentación**

En la primera sección de este capítulo, se presenta el esquema del algoritmo de segmentación automática que se propuso a utilizar. En la **sección 5.2,** se habla con detalle de otros algoritmos de pre-procesamiento propuestos con el fin de mejorar los resultados el algoritmo.

En la **sección 5.3** se discute detenidamente, sobre el esquema de búsqueda de rejilla para la validación cruzada y cuáles son los distintos hiperparametros a optimizar. Adicionalmente, también se discuten los diferentes casos de pruebas, mediante los cuales se evaluara el algoritmo. Mientras que, en la **sección 5.4**, se hablara de las diferentes métricas con las que se evaluara en algoritmos, además que se introducirán las herramientas que se utilizaran para poder facilitar su comprensión.

Finalmente, la **sección 5.5** es donde se analizan las etapas de validación y de evaluación para los diferentes casos de prueba establecidos en la **sección 5.3**. La **sección 5.5** se encuentra dividida en cuatro sub-secciones: las primeras tres sirven analizar individualmente cada uno de los casos de prueba establecidos. Mientras que, la **sección 5.5.4** recapitula cual ha sido la mejor de dichas implementaciones y en qué aspectos todavía existen mejoras.

## 5.1 Modelo propuesto

Para entrenar el algoritmo de inteligencia artificial se propone el modelo de la figura 5.1. Este consta de una etapa de pre-procesamiento de datos al donde se aplica la estandarización, métodos aproximativos de kernel y reducción de dimensiones usando PCA. Es necesario destacar que los métodos de Kernel son aplicados de forma explícita y no forman parte directa del algoritmo de SGD.

En la etapa de aprendizaje, se utiliza como algoritmo principal el SGD y los datos se preparan en un modelo de validación cruzada. Donde se ejecutara el proceso de la validación un total de tres veces para optimizar cada uno de los tres puntajes establecidos en la **sección 5.4.**

El modelo con el SGD optimizado es utilizado para ser probado con los datos de test y realizar los reportes de clasificación correspondientes, mientras que por otro lado, se puede utilizar el modelo optimizado para visualizar la predicción con una imagen original del dataset BRATS2017.

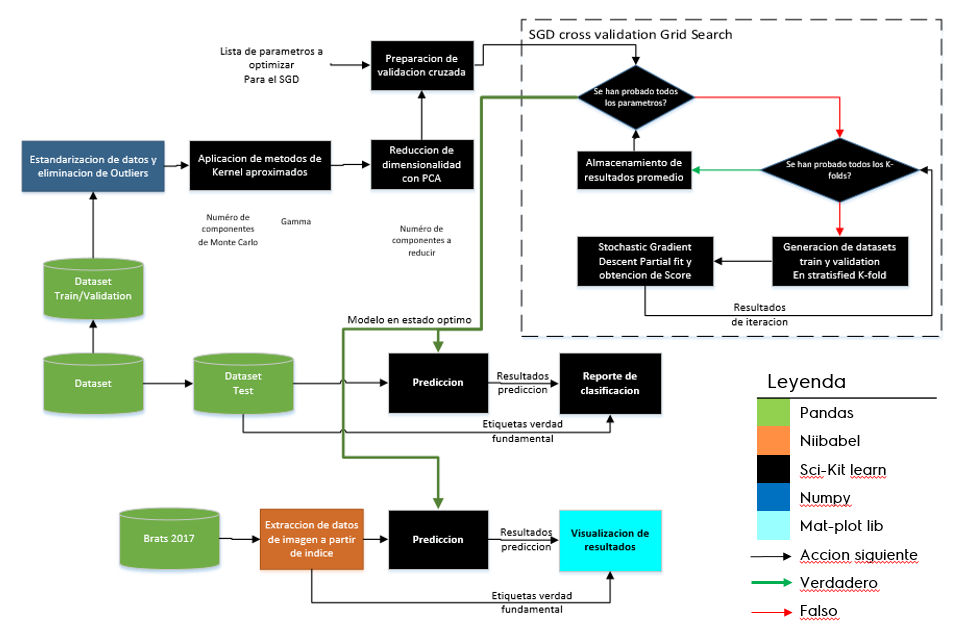
****

Ilustración 5.1 Esquema propuesto de modelo de aprendizaje

Los datos a escoger están divididos en dos conjuntos diferentes; uno de entrenamiento y uno de prueba. En el caso que se desee partir del dataset no reducido la distribución de los datasets de entrenamiento será de 180 pacientes y el dataset de prueba será de 30 pacientes. En el caso del dataset ya reducido a partir de lo que se propuso en el **capítulo 4** la proporción a utilizar será de 66% entre los datos de entrenamiento y de 33% los de prueba, respectivamente.

## 5.2 Aplicación de Kernel aproximados y PCA

Para el pre procesamiento de los datos se aplican métodos de Kernel aproximados y PCA para alterar el espacio de características de los pixeles y, de dicho modo, hallar una separación lineal entre las diferentes clases.

A pesar de que en la mayoría de los casos en que se utilizan algoritmos del tipo SVM los Kernel son calculados implícitamente dentro de dicho algoritmo. Para el caso del SGD es necesario calcular el mapeo de las muestras en el nuevo espacio de forma explícita para que sirva de entrada al algoritmo SGD. La librería de *sci-kit learn* utiliza el método de aproximación de Monte Carlo para hallar la aproximación al Kernel RBF. [46, 47, 48]

Para estos casos, existe una serie de hiperparametros adicionales a tener que optimizar, por ejemplo:

* Las **dimensiones** de la estimación de Kernel. Se tiene que para mientras mayor la cantidad de características, más precisa seria la implementación del Kernel no aproximado. Sin embargo, mientras mayor sea la precisión también se tendrá un procesamiento más lento de los datos debido a que se tienen mayores características por muestra.
* **Gamma** del Kernel RBF que se define como el valor γ por el cual se multiplica la función de Kernel y puede afectar la forma directa el mapeo realizado mediante dicha función.

Se aplica el PCA para reducir las dimensiones de la aproximación de Kernel implementado en el paso anterior. Así se puede obtener una reducción de datos y aprovechar de la ventaja del Kernel al cambiar la estructura de los datos para proponer un problema linealmente resoluble, en el caso del SGD.

## 5.3 Implementación del esquema de aprendizaje

La búsqueda en cuadricula usando validación cruzada, o *grid search cross validation* en inglés, permite obtener los parámetros óptimos, a partir de la ejecución de una serie de validaciones cruzadas [49]. Se aplicó la implementación disponible en *sci-kit learn* de la búsqueda en cuadricula [48]. Esta etapa es ejecutada tres veces, de acuerdo a los diferentes entradas que se usaran en la **sección 5.5** para su respectiva evaluación:

1. **Datos no reducidos:** para este caso, la entrada del algoritmo es el conjunto de datos cuando, **se omite** el algoritmo de reducción de datos establecido de las **secciones 4.6 a la 4.9**. Es decir, para esta entrada lo único que se aplica es la estandarización de los datos y la eliminación de los outliers, solamente con el fin de estabilizar el algoritmo del SGD.
2. **Datos reducidos:** Se aplica todo lo que se ha expuesto en el **capítulo 4**. Es decir, se aplica al conjunto de datos original la reducción de clústeres redundantes y la reducción de datos redundantes dentro de cada clústeres. Esta reducción se aplica con las modificaciones propuestas al algoritmo de Shen, ya que, como se estableció en la **seccion 4.10,** dichas modificaciones resultan más eficientes en cuanto a volumen de datos reducidos.
3. **Datos reducidos tras la aplicación de métodos de Kernel aproximados y PCA:** En este caso, la entrada aplica la reducción del **capítulo 4** y se aplican los pasos establecidos en la **sección 5.2** para mejorar los resultados del algoritmo.

Es importante destacar que cada una de estas entradas representa un caso de prueba diferente.

Como estrategia de *cross-validation*, se propone utilizar 5 *k-folds* estratificados para cada iteración dentro de la rejilla. De esta manera, se estima siempre tener muestras de ambas clases para cada *k-fold*. La rejilla ha sido definida de la siguiente manera:

Donde cada uno de los mencionados hiperparametros representan:

* **α** representa la constante que multiplica el termino de regularización. El valor de este parámetro afecta directamente la velocidad del algoritmo en converger a un resultado.
* El **número máximo de iteraciones**, representa la máxima cantidad de ejecuciones del SGD antes de converger a un resultado.
* El **balanceo de las muestras** permite asignar un peso a cada clase. Se propone como balance de cada muestra la inversa de la frecuencia de sus datos (en el caso de ser aplicado). Es decir:

Dadas las distintas permutaciones de los hiperparametros a optimizar, durante la búsqueda de rejilla se va a implementar un total de 12 ciclos de cross-validation, que a su vez implican 5 k-folds cada uno. Por cada iteración del K-fold se prevee un entrenamiento parcial y validación del SGD.

Además, cada uno de los 12 ciclos propone otorga un puntaje de comparar el entrenamiento, con respecto al conjunto de datos de validación perteneciente al ciclo, estos puntajes serán analizados con mejor detalle, en la **sección 5.5.**

El hecho de aplicar este tipo de croos-validation permite reducir el overfitting de un modelo, obteniendo mejores resultados que al aplicar solamente un train-test Split. Como el algoritmo a optimizar será un SGD, es un requisito indispensable estandarizar los datos antes de ejecutarlos como se ha discutido en la **sección 4.5**.

Tras esta etapa de cross-validation se espera obtener los mejores hiperparametros para el algoritmo de SGD. Sin embargo, para el tercer caso de prueba, donde se aplican los métodos de kernel aproximado y a la reducción de PCA, al ser etapas que se ejecutan explícitamente tendrán que ser obtenidos fuera de los métodos de evaluación que el algoritmo de cross-validation propone.

## 5.4 Métricas de clasificación

### 5.4.1 Matriz de confusión

La matriz de confusión [42, 43] es una tabla usada para describir el rendimiento de modelos de clasificación en un set de datos de test en el cual se conocen los verdaderos valores.

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | Clase predicha | |
| Clase  Real |  | Clase = No | Clase = Si |
| Clase = No | Negativo Verdadero | Falso Positivo |
| Clase = Si | Falso Negativo | Positivo Verdadero |

Tabla 5.1 Matriz de confusion

Los positivos y negativos reales son las observaciones que son predichas correctamente y por eso son mostradas en color verde. Se desea minimizar los falsos positivos y los falsos negativos que se muestran en color rojo.

* **Positivos Verdaderos (TP):** Son los términos correctamente predichos como positivos de tejido tumoral.
* **Negativos Verdaderos (TN)**: Son los datos correctamente predichos como negativos de tejido tumoral.
* **Falso Negativo (FN):** Cuando un resultado que en verdad es positivo es predicho como uno negativo.
* **Falso positivo (FP):** Resultados negativos predichos como positivos.

### 5.4.2 Reporte de clasificación

Las diferentes métricas de un reporte clasificación [44] nacen de la matriz de confusión y permiten entender la relación entre los falsos negativos, falsos positivos, los verdaderos positivos y verdaderos negativos y evaluar el desempeño de una clasificación. El reporte de clasificación utilizado se basa en aquel que aporta la biblioteca de sci-kit learn de Python. [45]

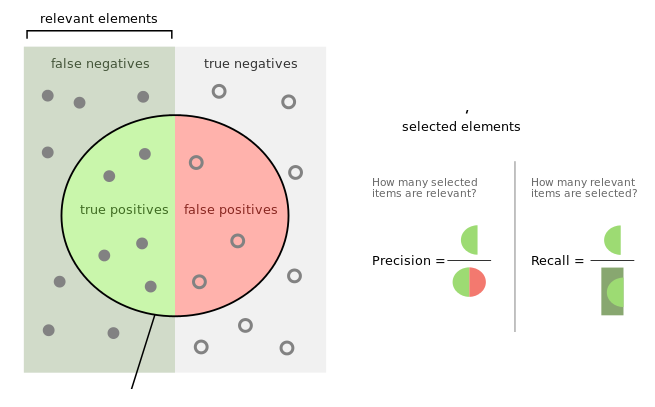


Ilustración 5.2 ilustracion de diferentes metricas de reporte de aprendizaje

**Accuracy (Rendimiento):** Es la forma más intuitiva de medir el rendimiento, y es la tasa entre las muestras correctamente predichas y las observaciones totales. Para medir el rendimiento global de una muestra se calcula de la siguiente forma:

**Precisión:** para una clase, es la tasa de las muestras correctamente predichas entre todas las observaciones que se predijeron como pertenecientes a dicha clase. Es decir:

Para las muestras positivas:

Para las muestras negativas:

**Recall (Sensibilidad):** para una clase, se define como la tasa de muestras que han sido clasificadas correctamente entre el conjunto de muestras de la misma clase que han sido clasificadas correctamente y las muestras que ha sido clasificadas como de otra clase. En otras palabras:

Para muestras positivas:

Para muestras negativas:

**F1-score:** es el promedio entre la sensibilidad y la precisión, evaluando los costos entre ambas clases. Es una tasa más efectiva de clasificación para muestras desbalanceadas entre la cantidad de clases. Se define del siguiente modo para cada clase:

Para muestras positivas:

Para muestras negativas:

**Support:** La cantidad de muestras para cada clase, aunque este valor no habla directamente del rendimiento de los resultados. Permite tener una idea del equilibrio que se puede observar entre las diferentes clases existentes.

## 5.5 Resultados y análisis

. Se presentan tres diferentes casos de pruebas, los cuales dependen de las entradas utilizadas en la **sección 5.3** (Datos no reducidos, datos reducidos y datos reducidos con métodos de aproximación de Kernel y PCA). Para cada uno de estos casos, se analizan los resultados mediante dos etapas diferentes tal y como se observa en la figura 5.3:

* **Fase de desarrollo:** Se analiza el desempeño de la validación cruzada mediante el método de búsqueda de rejilla. En este caso, el enfoque es observar cuales son los mejores hiperparametros para optimizar cada uno de los 3 puntajes definidos en la **sección 5.4.2.** Por lo tanto, se debería tener como máximo 3 modelos diferentes, donde cada uno, ha sido aquel con mejor resultado para optimizar uno de los tres puntajes.
* **Fase de evaluación:** Para esta fase, se procede a evaluar las predicciones de un conjunto de datos completamente desconocido para el algoritmo. Se analiza el desempeño del algoritmo dependiendo de los parámetros establecido en la Fase de desarrollo. Esta etapa se divide en dos sub-etapas:
  + **Fase de evaluación cuantitativa:** Se realizan reportes de clasificación y matrices de confusión para observar de manera cuantitativa como
  + **Fase de evaluación visual o cualitativa:** Se comparan los resultados con 4 pacientes determinados aleatoriamente (paciente 150, paciente 185, paciente 186 y paciente 205) y se visualizan para observar cual es el resultado con los tres modelos.

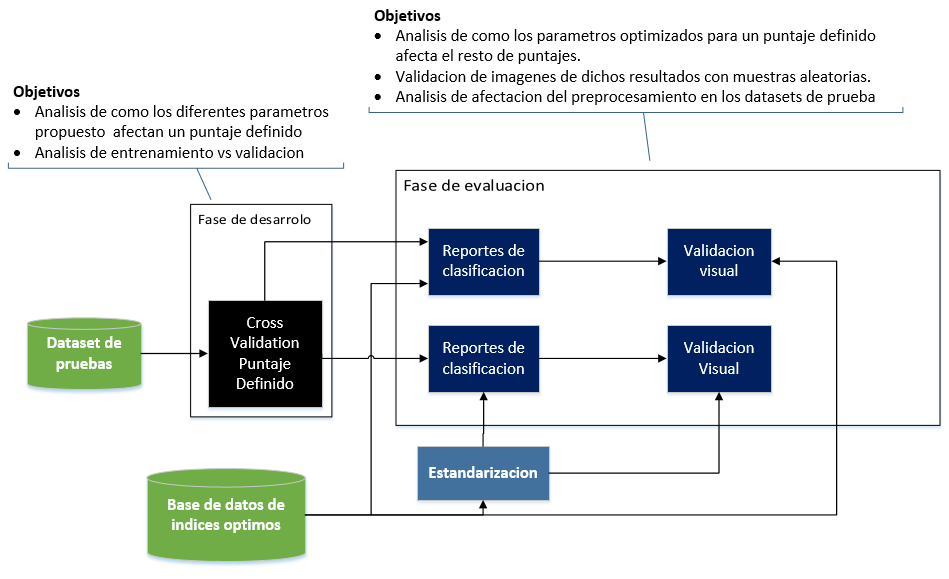


Ilustración 5.3 Esquema propuesto de evaluacion para un solo puntaje definido.

### 5.5.1 Análisis de rendimiento de SGD para los datos no reducidos.

Se presentan los resultados de entrenar el algoritmo de entrenamiento partiendo del conjunto de datos no reducidos. Cabe recordar, que para este conjunto, las imágenes óptimas extraídas (según el procedimiento explicado en la sección 3.4) de los primeros 180 pacientes del dataset global de BRATS\_2017 representaran las imágenes de entrenamiento y validación del algoritmo de SGD, mientras que los datos óptimos de los últimos 30 pacientes representan el conjunto de datos de prueba. Se evaluaran los mejores desempeños para la precisión, sensibilidad y el puntaje F1.

Para la etapa de evaluación es decir, el conjunto de datos de prueba. Se observara como el proceso de estandarización de datos sobre el conjunto de datos de prueba puede alterar estos resultados. Cabe destacar que no se aplica la eliminación de *outliers* para el conjunto de datos pruebas, debido a que, a pesar de ser valores muy alejados de las desviaciones estándar, es necesario comprobar que el algoritmo entrenado sea robusto contra los valores atípicos.

Adicionalmente, se observara la visualización de los resultados obtenidos para un paciente cuyos datos ya pertenecen al conjunto de datos de entrenamiento (paciente 150) y tres pacientes cuyos datos pertenecen al conjunto de datos de prueba (paciente 185, 186 y 205), seleccionados aleatoriamente. Así se estima, hallar cual es la mejor métrica a optimizar en la fase de desarrollo.

#### 5.5.1.1 Fase de desarrollo

Los resultados señalados en la fase de desarrollo representan el promedio y la desviación estándar (que será tomada como error) de las 5 iteraciones de los k-folds del SGD utilizando la validación cruzada. Dichos resultados son producto de la comparación entre los diferentes datos de entrenamiento y de datos de validación.

##### 5.5.1.1.1 Optimizando de la precisión

El conjunto de parámetros que optimiza la precisión del algoritmo de clasificación es aquel con una constante α de 0.001, un número máximo de iteraciones de 5 y al no balancear las clases. Para una precisión de 0.804 y una desviación estándar de ± 0.137.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Precision | Desv. Standard (±) | Alfa | Iteraciones | Balance |
| 0.804 | 0.137 | 0.001 | 5 | None |
| 0.804 | 0.136 | 0.001 | 10 | None |
| 0.803 | 0.138 | 0.001 | 20 | None |
| 0.686 | 0.077 | 0.001 | 5 | balanced |
| 0.685 | 0.076 | 0.001 | 10 | balanced |
| 0.685 | 0.076 | 0.001 | 20 | balanced |
| 0.799 | 0.137 | 0.0001 | 5 | None |
| 0.798 | 0.137 | 0.0001 | 10 | None |
| 0.801 | 0.134 | 0.0001 | 20 | None |
| 0.685 | 0.075 | 0.0001 | 5 | balanced |
| 0.685 | 0.075 | 0.0001 | 10 | balanced |
| 0.685 | 0.075 | 0.0001 | 20 | balanced |

Tabla 5.2 Promedio y error de los resultados de precisión de la validación cruzada para los diferentes parámetros establecidos en la búsqueda de rejilla con el dataset no reducido.

Se puede destacar que, globalmente, se obtienen mejores resultados al no balancear las muestras y tras un corto número de iteraciones. El factor tiene una influencia notoria en el desempeño del algoritmo y como es de esperarse, mientras menor sea el factor α mayor es la cantidad de iteraciones necesarias para alcanzar un mejor resultado.

La desviación estándar de los resultados de precisión hace que se puedan considerar esta implementación como inestable, considerando que para muestras no balanceadas es de, en promedio, ±0.137 y aun así balanceando las muestras se tiene un promedio de ±0.075.

##### 5.5.1.1.2 Optimización de la sensibilidad

El conjunto de parámetros que optimiza la sensibilidad del algoritmo de clasificación es aquel con una constante α de 0.0001, un número máximo de iteraciones de 20 y un balanceo en las muestras según la proporción de las clases. Para una sensibilidad máxima de 0.776 y un error de ± 0.094

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Sensibilidad | Desv. Standard (±) | Alfa | Iteraciones | Balance |
| 0.629 | 0.057 | 0.001 | 5 | None |
| 0.629 | 0.054 | 0.001 | 10 | None |
| 0.630 | 0.057 | 0.001 | 20 | None |
| 0.775 | 0.095 | 0.001 | 5 | balanced |
| 0.775 | 0.094 | 0.001 | 10 | balanced |
| 0.775 | 0.095 | 0.001 | 20 | balanced |
| 0.635 | 0.063 | 0.0001 | 5 | None |
| 0.634 | 0.059 | 0.0001 | 10 | None |
| 0.633 | 0.056 | 0.0001 | 20 | None |
| 0.775 | 0.094 | 0.0001 | 5 | balanced |
| 0.775 | 0.094 | 0.0001 | 10 | balanced |
| 0.776 | 0.093 | 0.0001 | 20 | balanced |

Tabla 5.3 Promedio y error de los resultados de sensibilidad de la validación cruzada para los diferentes parámetros establecidos en la búsqueda de rejilla con el dataset no reducido.

Se observó que, en términos de sensibilidad, se toma como el mejor modelo aquel donde se estima el balance propuesto. Sin embargo, se requiere aparentemente de un mayor número de iteraciones en esta configuración para alcanzar una convergencia.

Se tiene que para los diferentes conjuntos de parámetros de sensibilidad existe una desviación estándar de en promedio ±0.057 para las muestras no balanceadas y ±0.097 para las muestras balanceadas. Esto implica que, para diferentes conjuntos de muestra reducidas se pueden conseguir resultados inestables y no se pueden garantizar que los resultados no varíen si se cambia el conjunto de evaluación.

##### 5.5.1.1.3 Optimización del puntaje F-1

Se obtuvo que el mejor modelo para este caso se obtiene al balancear los datos, utilizar un factor α de 0.0001 y un numero de 5 iteraciones. Para un resultado de, en promedio, 0.702 ± 0.092.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| F1-score | Desv. Standard (±) | Alfa | Iteraciones | Balance |
| 0.663 | 0.073 | 0.001 | 5 | None |
| 0.663 | 0.073 | 0.001 | 10 | None |
| 0.662 | 0.071 | 0.001 | 20 | None |
| 0.701 | 0.096 | 0.001 | 5 | balanced |
| 0.701 | 0.097 | 0.001 | 10 | balanced |
| 0.701 | 0.096 | 0.001 | 20 | balanced |
| 0.669 | 0.074 | 0.0001 | 5 | None |
| 0.666 | 0.074 | 0.0001 | 10 | None |
| 0.668 | 0.074 | 0.0001 | 20 | None |
| 0.702 | 0.092 | 0.0001 | 5 | balanced |
| 0.701 | 0.093 | 0.0001 | 10 | balanced |
| 0.700 | 0.097 | 0.0001 | 20 | balanced |

Tabla 5.4 Promedio y error de los resultados de f1score de la validación cruzada para los diferentes parámetros establecidos en la búsqueda de rejilla con el dataset no reducido.

Se observa que, al igual que ocurre con la optimización de sensibilidad, los mejores resultados son aquellos donde se plantea un balance entre las muestras, sin embargo a diferencia de la optimización anterior, se requiere de un menor número de iteraciones para obtener mejores resultados.

De la manera anterior, se observa que estos parámetros también representa valores de desviación estándar superiores a ±0.05, lo que implica que se tiene poca fiabilidad para esta métrica para muestras desconocidas. Considerando que los valores del puntaje F1 son calculados en base a la precisión y sensibilidad del algoritmo. En este dataset, se puede inferir que la sensibilidad influye más en el puntaje F1 que la precisión de los resultados del puntaje F1.

#### 5.5.1.2 Fase de evaluación

En esta fase se comparara el rendimiento de los tres diferentes modelos entrenados para optimizar la precisión, la sensibilidad y el puntaje F1, con respecto al mismo conjunto de datos de prueba, antes y después de estandarizarlos. Además, se analizara como el efecto de optimizar un puntaje puede impactar el rendimiento de los otros. Por ejemplo, al optimizar la precisión como se vería afectada la sensibilidad y el puntaje F1.

Al evaluar el **modelo que optimiza la precisión**, se detalla que:

* Se obtienen globalmente mejores resultados al no estandarizar los datos de prueba en las tres puntaciones.
* Sin embargo, observando individualmente los resultados para la clase 0 (tejido no tumoral), se observa que hay una recaída de 0.05 en la sensibilidad, considerando que, dado que el volumen de datos de clase 0 es 3.9 veces mayor a los de clase 1 se puede inducir que una reducción de 0.05 implica que existirá un mayor número de falsos positivos en la clasificación tras estandarizar el conjunto de datos de prueba.
* Por otro lado, observando individualmente la clasificación para la clase 1 (tejido tumoral), se puede observar que la sensibilidad mejora al no estandarizar el conjunto de datos de prueba. Además, el hecho de que la sensibilidad aumente para la clase 1 tras la estandarización, mientras la precisión permanezca igual, puede implicar una reducción de falsos negativos.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Precision | | recall | | F1-score | | support |
| Clase | Estand | No  Estand | Estand | No  Estand | Estand | No  Estand |
| 0 | 0.84 | 0.88 | 0.95 | 0.90 | 0.89 | 0.89 | 967144 |
| 1 | 0.60 | 0.57 | 0.27 | 0.53 | 0.37 | 0.55 | 247188 |
| avg / tot | 0.79 | 0.82 | 0.81 | 0.82 | 0.79 | 0.82 | 1214332 |

Tabla 5.5 Reporte de clasificación de los resultados de precisión para el dataset no reducido.

En efecto, al comparar el reporte de clasificación con la matriz de confusión, se puede observar que estandarizar el conjunto de datos de prueba aumenta la tendencia del modelo a indicar que un pixel pertenece a la clase tumoral. Por ejemplo, en la matriz de confusión se evidencia un aumento de los verdaderos positivos y de los falsos positivos.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  | Clase predicha | | | |
|  |  | Test estandarizado | | Test no estandarizado | |
|  |  | No Tumoral | Tumoral | No tumoral | Tumoral |
| Clase  Real | No tumoral | 921 740 | 45 404 | 867 414 | 99 730 |
| Tumoral | 180 313 | 66 875 | 116 781 | 130 407 |

Tabla 5.6 Matriz de confusión de los resultados de precisión para el dataset no reducido.

Respecto a la **optimización de la sensibilidad**, se observa que:

* globalmente la sensibilidad disminuye en 0.7 al estandarizar los datos. Sin embargo, a nivel global la precisión permanece en niveles similares al solo reducir 0.02 tras la estandarización.
* Los valores de precisión para la clase de tejido tumoral disminuyen hasta 0.19 con respecto al modelo anterior, mientras que la estandarización afecta negativamente a los resultados de la precisión. Sin embargo, los parámetros de sensibilidad aumentan en 0.04, lo cual se traduciría en mayor cantidad de falsos positivos tras la estandarización.
* Para al tejido no tumoral no existe variación de la precisión. Mientras la estandarización la varía en 0.09 la sensibilidad y, como consecuente, el puntaje F1 también se ve reducido en 0.06.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Précision | | Recall | | f1-score | | support |
| Clase | Estand | No  Estand | Estand | No  Estand | Estand | No  Estand |
| 0 | 0.90 | 0.90 | 0.76 | 0.85 | 0.82 | 0.88 | 967144 |
| 1 | 0.41 | 0.52 | 0.67 | 0.63 | 0.51 | 0.57 | 247188 |
| avg / tot | 0.80 | 0.82 | 0.74 | 0.81 | 0.76 | 0.81 | 1214332 |

Tabla 5.7 Reporte de clasificación de los resultados de Sensibilidad para el dataset no reducido.

Estos parámetros se traducen en que la estandarización tiene un efecto negativo en los resultados de las imágenes al aumentar la tasa de falsos positivos en 39.15% mientras que los valores de falsos negativos y verdaderos positivos han variado solamente en 10.77% y 6.44%, respectivamente.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  | Clase predicha | | | |
|  |  | Test estandarizado | | Test no estandarizado | |
|  |  | No Tumoral | Tumoral | No tumoral | Tumoral |
| Clase  Real | No tumoral | 733 387 | 233 757 | 824 892 | 142 252 |
| Tumoral | 82 611 | 1. 577 | 92581 | 154 607 |

Tabla 5.8 Matriz de confusión de los resultados de sensibilidad para el dataset no reducido.

El **modelo que optimiza el puntaje F1** demostró ser similar durante la fase de desarrollo al modelo que optimiza la sensibilidad como se estimó en la fase desarrollo de este modelo, se observa en el reporte de clasificación que el único verdadero cambio es el valor global de la precisión al no estandarizar los datos.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Precision | | Recall | | f1-score | | support |
| Clase | Estand | No  Estand | Estand | No  Estand | Estand | No  Estand |
| 0 | 0.90 | 0.90 | 0.76 | 0.85 | 0.82 | 0.87 | 967144 |
| 1 | 0.41 | 0.52 | 0.67 | 0.63 | 0.51 | 0.57 | 247188 |
| avg / tot | 0.80 | 0.82 | 0.74 | 0.80 | 0.76 | 0.81 | 1214332 |

Tabla 5.9 Reporte de clasificación de los resultados del puntaje F1 para el dataset no reducido.

Analizando la matriz de confusiones se tiene que el efecto es similar al del modelo anterior considerando que los falsos positivos varían en un 38.08%

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  | Clase predicha | | | |
|  |  | Test estandarizado | | Test no estandirado | |
|  |  | No Tumoral | Tumoral | No tumoral | Tumoral |
| Clase  Real | No tumoral | 730 296 | 236 848 | 820490 | 146654 |
| Tumoral | 81029 | 166 159 | 90751 | 156437 |

Tabla 5.10 Matriz de confusión de los resultados del puntaje F1 para el dataset no reducido.

Para la **validación visua**l, se utilizaron las siguientes imágenes para cada paciente estudiado (Comenzando de la esquina superior izquierda y terminando en la esquina inferior derecha):

1. Predicción a partir del modelo con la **precisión optima**, se utiliza como entrada los datos del paciente **estandarizados.**
2. Predicción a partir del modelo con la **precisión optima**, se utiliza como entrada los datos del paciente **no estandarizados.**
3. Predicción a partir del modelo con la **sensibilidad optima**, se utiliza como entrada los datos del paciente **estandarizados.**
4. Predicción a partir del modelo con la **sensibilidad optima**, se utiliza como entrada los datos del paciente **no estandarizados.**
5. Predicción a partir del modelo con el **puntaje F1 óptimo**, se utiliza como entrada los datos del paciente **estandarizados.**
6. Predicción a partir del modelo con el **puntaje F1 óptimo**, se utiliza como entrada los datos del paciente **no estandarizados.**
7. Resultados de la verdad fundamental.
8. Imagen original.

Observando la imagen del paciente 150 cuyos datos forman parte del conjunto de datos de entrenamiento (donde se estima una segmentación aceptable) se detalla lo siguiente:

* Las predicciones para los datos no estandarizados son similares entre ellas, para los tres modelos entrenados.
* Los modelos de sensibilidad y para el puntaje F1 tienen predicciones similares al ser modelos que solo difieren en la cantidad máxima de iteraciones.
* El modelo entrenado para obtener una mejor sensibilidad resulta más robusto al ruido y de aportar falsos positivos a la segmentación.



Ilustración 5.4 Validación visual con el paciente 150 y datos no reducidos.

Posteriormente analizando las imágenes de los otros tres pacientes cuyos datos forman parte del conjunto de datos de prueba (Es decir, son completamente desconocidos para el SGD) se detalla lo siguiente:

* El modelo de precisión tienen menor tendencia a generar falsos positivos. Sin embargo, no logra detectar la misma cantidad de tejido tumoral que los modelos para la sensibilidad y el puntaje F1 óptimos. Es decir que no logra detectar la masa tumoral completamente en todos los casos, sin embargo logra segmentar parte de ella.

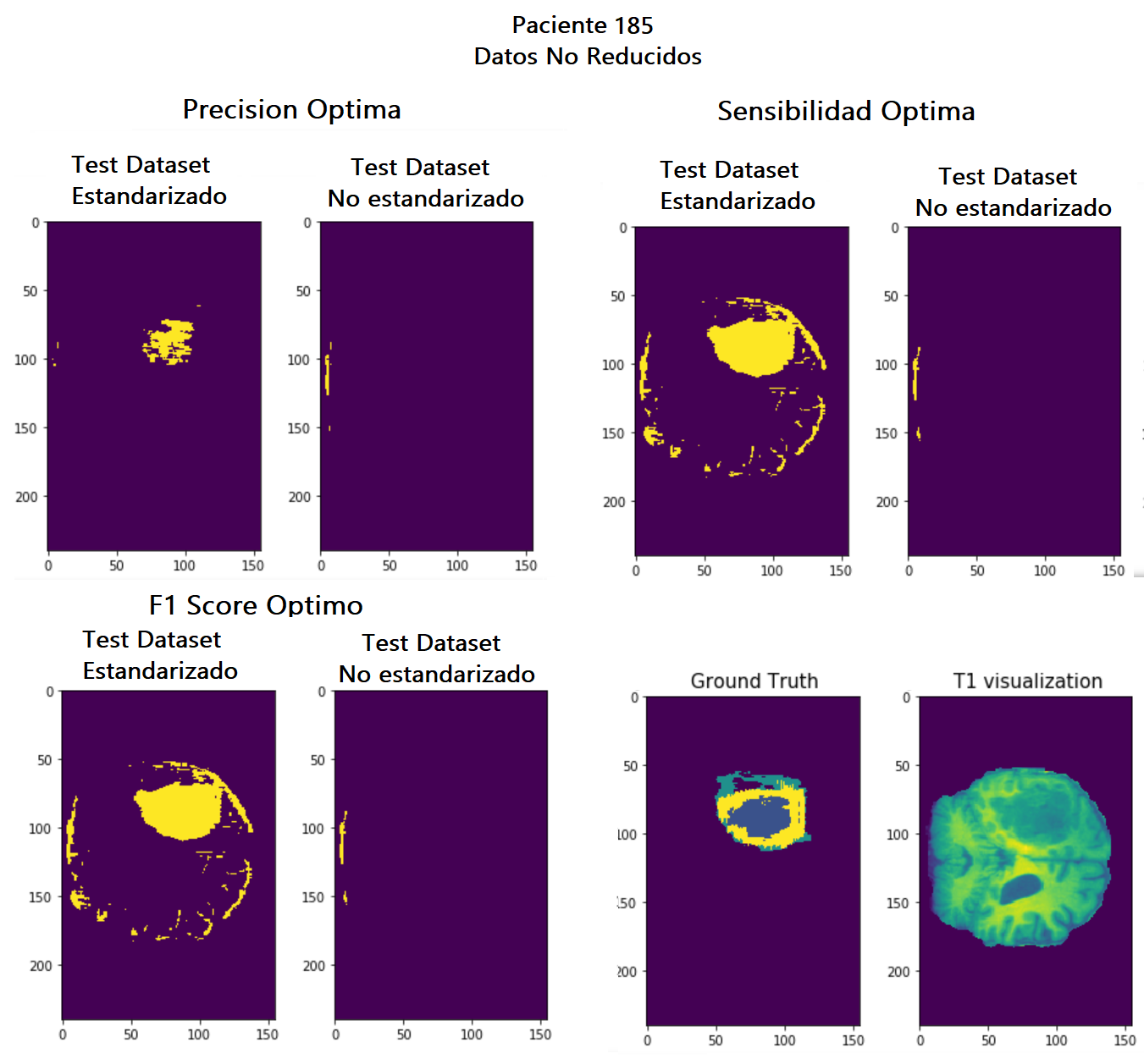


Ilustración 5.5 Validación visual con el paciente 185 y datos no reducidos.

* No hay variaciones significativas entre las optimizaciones de sensibilidad y puntaje F1 para ninguna de las imágenes de los tres pacientes.
* Los modelos para la sensibilidad y el puntaje F1 logran detectar más pixeles de la masa tumoral que el modelo de precisión. No obstante, en los casos del paciente 185 (Fig 5.5) y 205 (Fig 5.7) se aportan falsos positivos, sobre todo al borde del cerebro.
* Para los tres modelos que han sido entrenados (sensibilidad, precisión y puntaje F1) se observa que los resultados son todos similares cuando el dataset de prueba no está estandarizado.
* Todos los modelos responden con relativa facilidad al tejido necrótico del cerebro (En color azul). Sin embargo, los modelos no son completamente eficientes para la detección del edema, como se previó en la **sección 4.2.1**.

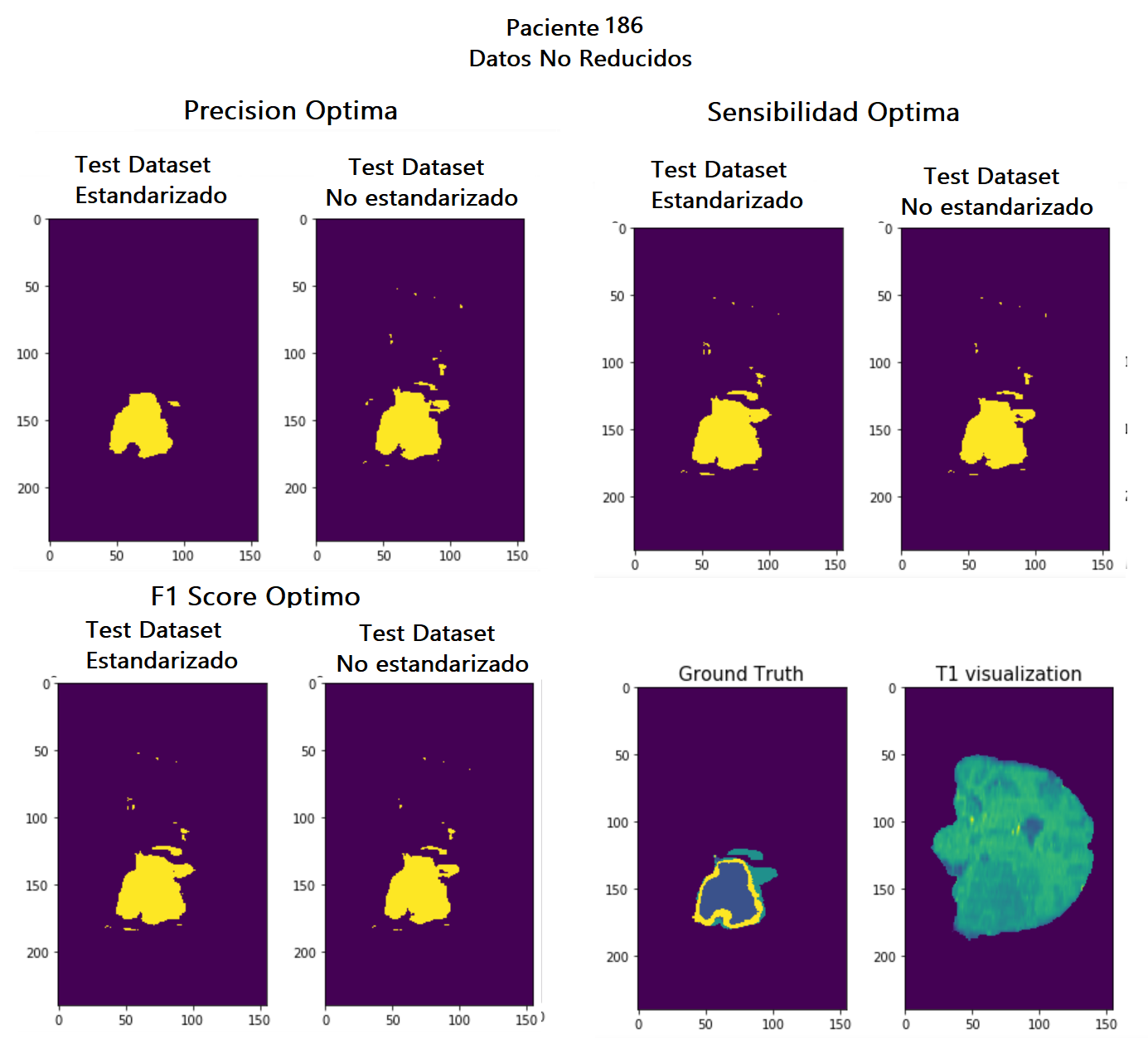


Ilustración 5.6 Validación visual con el paciente 186 y datos no reducidos.

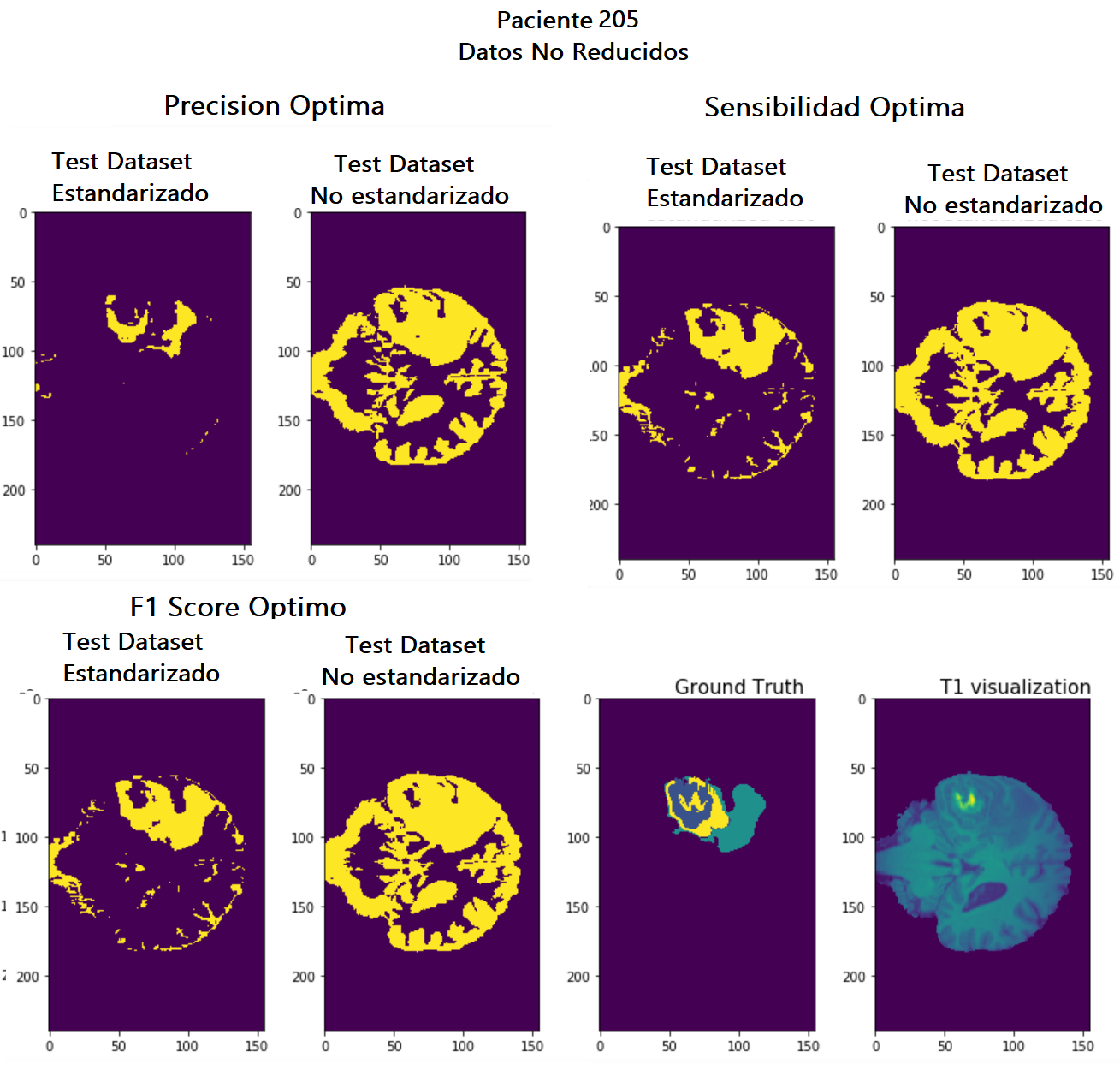


Ilustración 5.7 Validación visual con el paciente 205 y datos no reducidos.

### 5.5.2 Análisis de desempeño de SGD para los datos reducidos

De una manera similar al análisis de datos no reducidos, se dividen los resultados en la fase de desarrollo y la fase de evaluación como se explica en la imagen 5.3. Sin embargo, para cada etapa se contrastaran los resultados con respecto a los resultados de la **sección 5.5.1**, con el fin de determinar si la reducción de datos ha hecho el algoritmo más eficiente.

#### 5.5.2.1 Fase de desarrollo

La primera diferencia, de reducir los datos, se puede observar en el rendimiento de tiempo de desarrollo para cada uno de los datasets. Se midió el tiempo de las fases de desarrollo de la **sección 5.5.1.1** para cada una de las posibles optimizaciones de ambos conjuntos de datos y se comparó con los de esta sección. Los resultados son expuestos en la tabla 5.11.

|  |  |  |
| --- | --- | --- |
|  | Dataset | |
| Factor a optimizar | Reducido | No reducido |
| Precision | 6min 33s | 40min 36s |
| Recall | 9min 41s | 21min 21s |
| F1 | 6min 49s | 19min 19s |

Tabla 5.11 Tiempo de desarrolo de datos reducidos vs Datos No reducidos

Se detalla como el tiempo de procesamiento computacional ha sido reducido a un 16.1% respecto a la optimización de precisión; a un 30.7%, la optimización de los parámetros de precisión y; a un 43.9% la optimización de parámetros de F1. Por lo tanto, en términos de costo computacional resulta más eficiente el set de datos tras la reducción propuesta en el **capítulo 4**.

##### 5.5.2.1.1 Optimización de la precisión

En la siguiente tabla, se halla que el mejor modelo para la optimización de la precisión es aquel con un muestreo no balanceado, un número de 5 iteraciones y una constante α de 0.0001. Donde se obtuvo en promedio una precisión de 0.804 y un error de ± 0.015. Resultados en la tabla 5.12.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Precision | Desv. Standard (±) | Alfa | Iteraciones | Balance |
| 0.802 | 0.005 | 0.001 | 5 | None |
| 0.802 | 0.004 | 0.001 | 10 | None |
| 0.801 | 0.004 | 0.001 | 20 | None |
| 0.662 | 0.001 | 0.001 | 5 | balanced |
| 0.662 | 0.001 | 0.001 | 10 | balanced |
| 0.662 | 0.001 | 0.001 | 20 | balanced |
| 0.804 | 0.015 | 0.0001 | 5 | None |
| 0.800 | 0.014 | 0.0001 | 10 | None |
| 0.801 | 0.005 | 0.0001 | 20 | None |
| 0.662 | 0.002 | 0.0001 | 5 | balanced |
| 0.663 | 0.003 | 0.0001 | 10 | balanced |
| 0.662 | 0.002 | 0.0001 | 20 | balanced |

Tabla 5.12 Promedio y error de los resultados de precisión de la validación cruzada para los diferentes parámetros establecidos en la búsqueda de rejilla con el dataset reducido

El modelo es más preciso mientras no se tome un balance entre las diferentes muestras. Sin embargo, el modelo resulta tener mejores resultados cuando el factor *α* es inferior.

Cabe destacar, que los valores de precisión son similares al caso en el que se toman como set de datos de entrenamiento los datos no reducidos (0.804 ± 0.137). Sin embargo, el verdadero cambio es la desviación estándar, donde los índices de error son inferiores al caso mencionado. Por lo tanto, se considera que este modelo con reducción de datos puede ser más estable para futuras predicciones y menos sobre ajustado.

##### 5.5.2.1.2 Optimización de la sensibilidad

Respecto a la sensibilidad, el mejor modelo es aquel que tiene un α de 0.0001, un máximo de 10 iteraciones y un balanceo en las clases para alcanzar un puntaje de 0.727 ± 0.001. Véase la tabla 5.13.

Este valor es menor que cuando se usan los datos de entrenamiento no reducidos (0.776 ± 0.093). No obstante, el error es menor para los datos reducidos y se puede considerar que se trabaja con un modelo más robusto para futuras muestras desconocidas. Como en el caso anterior, el comportamiento de los resultados respecto a la sensibilidad de la clasificación es mejor cuando se realiza el balanceo entre las clases.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Sensibilidad | Desv. Standard (±) | Alfa | Iteraciones | Balance |
| 0.561 | 0.003 | 0.001 | 5 | None |
| 0.560 | 0.002 | 0.001 | 10 | None |
| 0.562 | 0.001 | 0.001 | 20 | None |
| 0.728 | 0.001 | 0.001 | 5 | Balanced |
| 0.728 | 0.001 | 0.001 | 10 | Balanced |
| 0.728 | 0.001 | 0.001 | 20 | Balanced |
| 0.577 | 0.029 | 0.0001 | 5 | None |
| 0.571 | 0.018 | 0.0001 | 10 | None |
| 0.566 | 0.014 | 0.0001 | 20 | None |
| 0.728 | 0.003 | 0.0001 | 5 | balanced |
| 0.727 | 0.001 | 0.0001 | 10 | balanced |
| 0.728 | 0.002 | 0.0001 | 20 | balanced |

Tabla 5.13 Promedio y error de los resultados de sensibilidad de la validación cruzada para los diferentes parámetros establecidos en la búsqueda de rejilla con el dataset reducido

##### 5.5.2.1.3 Optimización del puntaje F1

A partir de la tabla 5.14, los mejores parámetros para el puntaje F1 es un número de 20 iteraciones, una constante α de 0.0001 y un balance entre las diferentes muestras, para obtener en promedio un valor de 0.675 y un error de 0.002.

En el caso de las muestras no reducidas dicho valor alcanzaba, 0.702 ± 0.092. De nuevo, la reducción de datos provee un índice de error inferior.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| F1 - score | Desv. Standard (±) | Alfa | Iteraciones | Balance |
| 0.564 | 0.011 | 0.001 | 5 | None |
| 0.563 | 0.007 | 0.001 | 10 | None |
| 0.563 | 0.006 | 0.001 | 20 | None |
| 0.674 | 0.000 | 0.001 | 5 | balanced |
| 0.674 | 0.002 | 0.001 | 10 | balanced |
| 0.674 | 0.001 | 0.001 | 20 | balanced |
| 0.550 | 0.058 | 0.0001 | 5 | None |
| 0.570 | 0.019 | 0.0001 | 10 | None |
| 0.577 | 0.008 | 0.0001 | 20 | None |
| 0.674 | 0.006 | 0.0001 | 5 | balanced |
| 0.675 | 0.003 | 0.0001 | 10 | balanced |
| 0.675 | 0.002 | 0.0001 | 20 | balanced |

Tabla 5.14 Promedio y error de los resultados del puntaje F1 de la validación cruzada para los diferentes parámetros establecidos en la búsqueda de rejilla con el dataset reducido

#### 5.5.2.2 Fase de evaluación

Para la fase de evaluación de los datos reducidos se tomara un enfoque comparativo, respecto a la **sección 5.5.1.2**. Cabe destacar que, para esta etapa, el efecto de la estandarización de los datos sobre la evaluación de los datos queda fuera del perímetro de evaluación.

En primer lugar para el **modelo optimizado para la precisión** se observa:

* Los valores promedios de precisión se encuentran dentro del rango estimado durante la fase de desarrollo.
* Los valores globales son superiores al caso de muestras no reducidas en cuanto a precisión y sensibilidad. Sin embargo, se observa una caída en el puntaje F1 global.
* La precisión para el tejido tumoral es superior, a pesar de que exista una recaída en cuanto a la sensibilidad con respecto a los datos no reducidos.
* Respecto a la clase de tejido no tumoral se observa una mejora en la sensibilidad. Por lo tanto se estima una reducción de falsos negativos.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Clase | Precision | Sensibilidad | F1 | Support |
| 0 | 0.83 | 0.98 | 0.90 | 936999 |
| 1 | 0.74 | 0.21 | 0.33 | 235282 |
| Avg/tot | 0.81 | 0.83 | 0.79 | 1172281 |

Tabla 5.15 Reporte de clasificación de los resultados de precision para el dataset reducido

Analizando la matriz de confusión se observa que la tasa de verdaderos positivos es 57.93% menor que para el caso de muestras no reducidas (Ver tabla 5.6) y los verdaderos positivos representan un total del 46.05% de los verdaderos positivos. Sin embargo, analizando los falsos negativos, la tasa solo varia 7,4 % con respecto al caso de las muestras no reducidas tras la estandarización.

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | Clase predicha | |
|  |  | No tumoral | Tumoral |
| Clase  Real | No tumoral | 917900 | 19099 |
| Tumoral | 193813 | 41469 |

Tabla 5.16 Matriz de confusion de los resultados de precision para el dataset reducido

Con respecto a la **sensibilidad** se detalla lo siguiente:

* Los valores promediados para la precisión son similares al caso de datos no reducidos (Ver tabla 5.7). Sin embargo, los valores de sensibilidad y puntaje F1 son inferiores.
* Para la clase de tejido tumoral, los valores de precisión son inferiores y los de sensibilidad son superiores a los del caso mencionado.
* Para la clase de tejido sano, los valores de precisión son similares y; los de la sensibilidad, son inferiores. Esto aporta un mayor número de falsos negativos.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Clase | Precision | Sensibilidad | F1 | Support |
| 0 | 0.91 | 0.71 | 0.80 | 936999 |
| 1 | 0.39 | 0.73 | 0.51 | 235282 |
| avg / tot | 0.81 | 0.71 | 0.74 | 1172281 |

Tabla 5.17 Reporte de clasificación de los resultados de sensibilidad para el dataset reducido

En comparación con el caso anterior (ver tabla 5.8) se observa que la tasa de falsos negativos es superior, mientras se observa una reducción en la tasa de falsos positivos. Sin embargo, en este caso la proporción los falsos positivos es 1.59 veces la cantidad de verdaderos positivos, por lo tanto, existirán falsos volúmenes tumorales en las imágenes que superaran el tamaño del verdadero tejido tumoral.

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | Clase predicha | |
|  |  | No tumoral | Tumoral |
| Clase  Real | No tumoral | 661539 | 275460 |
| Tumoral | 62508 | 172774 |

Tabla 5.18 Matriz de confusion de los resultados de sensibilidad para el dataset reducido

El modelo **optimizado para el puntaje F1,** presenta las siguientes características:

* No las diferencias globales con respecto al modelo optimizador de la sensibilidad para el mismo conjunto de datos (Tabla 5.17). Considerando que la única diferencia entre los dos parámetros es la cantidad de iteraciones, los cambios respecto al funcionamiento del modelo no son significativos.
* Respecto al modelo de datos no reducidos (Tabla 5.9) los valores tienen un desempeño menor para la sensibilidad y el puntaje F1 promediados. Al reducir globalmente la sensibilidad y no variar la precisión se estima un aumento en predicciones erróneas.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Clase | Precision | Sensibilidad | Puntaje F1 | Support |
| 0 | 0.91 | 0.70 | 0.79 | 936999 |
| 1 | 0.38 | 0.74 | 0.50 | 235282 |
| avg / tot | 0.81 | 0.71 | 0.73 | 1172281 |

Tabla 5.19 Reporte de clasificación de los resultados del puntaje F1 para el dataset reducido

En la matriz de confusión se evidencia la misma desproporción entre los falsos positivos y verdaderos positivos que en la tabla . Se considera, que este modelo puede detectar masas tumorales donde menos del 50% será realmente tejido tumoral.

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | Clase predicha | |
|  |  | No tumoral | Tumoral |
| Clase  Real | No tumoral | 655630 | 281369 |
| Tumoral | 61108 | 174174 |

Tabla 5.20 Matriz de confusion de los resultados del puntaje F1 para el dataset reducido

Se recuerda que, para el **análisis de las imágenes** del conjunto de datos reducido, se descartó la visualización de las imágenes no estandarizadas, ya que, como se determinó en la sección anterior, estas imágenes no se adaptan al modelo que ha sido entrenado inicialmente y los resultados no son aceptables.

Para el paciente 150, cuyos datos forman parte del set de datos de entrenamiento, se puede observar que el mejor resultado ha sido aquel del modelo de precisión. Por otro lado, los modelos de sensibilidad y puntaje F1 no representan una variabilidad evidente. Además, ninguno de los dos logra segmentar el tejido tumoral de forma aceptable, ya que estos segmentan erróneamente un gran volumen del tejido sano como ‘tumor’.

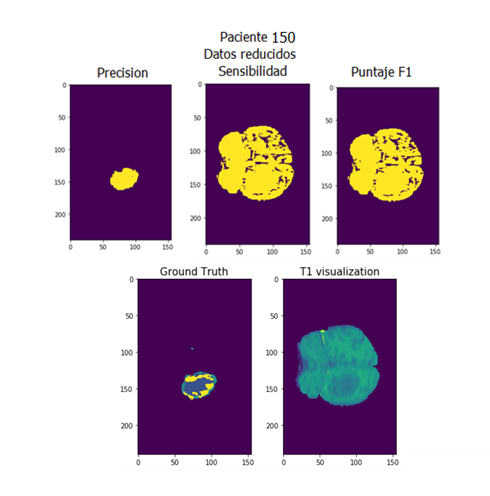


Ilustración 5.8 Validación visual paciente 150, en caso de datos reducidos

Respecto a las imágenes de pacientes 185, 186 y 205 (aquellos cuyos datos no pertenecen al conjunto de entrenamiento) se puede observar:

* Los **modelos optimizados para la sensibilidad y puntaje F1** aportan resultados muy similares entre sí. Sin embargo, ninguno de estos resultados permite tener una segmentación aceptable del tejido tumoral debido a que la cantidad de falsos positivos es mayor a la de los verdaderos positivos, como ha sido estimado a partir de las matrices de confusión.
* El **modelo de sensibilidad de la sección 5.5.1**, en comparación a su homólogo actual**,** ha sido capaz de detectar mejor los tejidos tumorales y de mantener los mismos niveles de precisión al no aportar falsos positivos al estudio como en donde la reducción es aplicada.

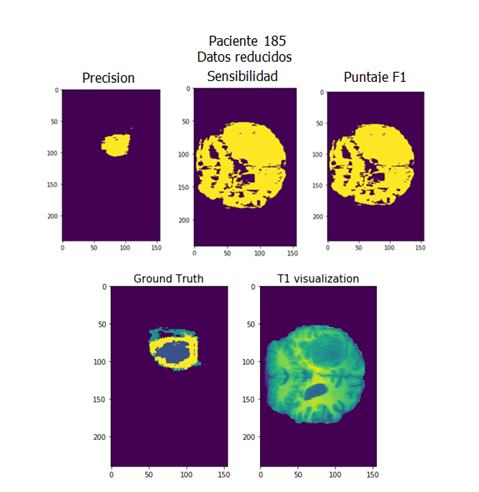


Ilustración 5.9 Validacion visual paciente 185, en caso de datos reducidos

* Las segmentaciones de precisión para esta sección son más sensibles a la detección del tejido necrótico que para los edemas.

De esta forma, se puede indicar que el conjunto de datos reducidos son más efectivos que los conjuntos de la sección anterior para el modelo de precisión, al lograr reducir el ruido. Sin embargo, el modelo no realiza predicciones aceptables si se optimiza la sensibilidad o el puntaje F1.

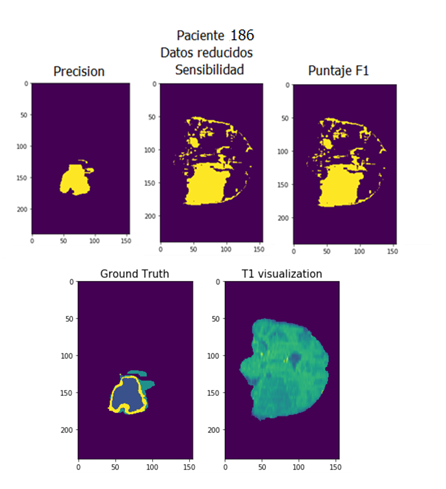


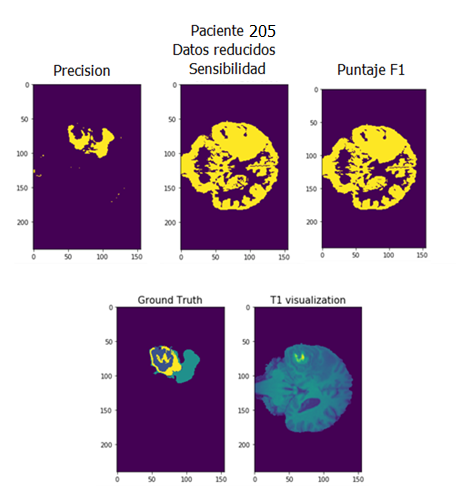
Ilustración 5.10 Validación visual paciente 186, en caso de datos reducidos  


Ilustración 5.11 Validación visual 205, en caso de datos reducidos

### 5.5.3 Análisis de desempeño para los métodos de Kernel

En las **secciones 5.5.1 y 5.5.2** se discutió el desempeño del SGD en el caso de aplicar o no la reducción de datos establecida en el **capítulo 4**. Entonces, se concluyó, al final de la **sección 5.5.2,** que aplicar la reducción de datos resulta más eficiente. Por lo tanto, en esta sección se estudia el desempeño del conjunto de datos al aplicar métodos aproximados de Kernel, se tomara como base el conjunto de datos reducidos.

#### 5.5.3.1 Fase de desarrollo

Como se plantea utilizar el conjunto de datos reducidos, se plantea optimizar solamente el SGD para conseguir los mejores parámetros de precisión. Ya que, como se determinó en la **sección 5.5.2**, optimizar los valores de sensibilidad y puntaje F1 a partir del modelo de datos reducidos no otorga resultados eficientes.

Es importante resaltar que, al tener que calcular explícitamente las aproximaciones de Kernel, es necesario considerar cada variación en los hiperparametros de los métodos de Kernel como un conjunto de datos completamente independiente al resto. Entonces, como la validación cruzada solo puede ser realizada para un mismo conjunto de datos al tiempo, se resaltara en la siguiente tabla los valores óptimos para cada cambio en los hiperparametros del Kernel.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Resultados | | Parametros Kernel | | Parametros SGD | | |
| Promedio | Desv standard ± | N | Gamma | Balance | Alpha | Max iter |
| 0.818 | 0.005 | 50 | 0.15 | None | 0.001 | 5 |
| 0.819 | 0.003 | 50 | 0.15 | None | 0.001 | 10 |
| 0.677 | 0.003 | 50 | 0.15 | Balanced | 0.001 | 5 |
| 0.676 | 0.003 | 50 | 0.15 | Balanced | 0.001 | 10 |
| 0.818 | 0.005 | 50 | 0.1 | None | 0.001 | 5 |
| 0.819 | 0.004 | 50 | 0.1 | None | 0.001 | 10 |
| 0.675 | 0.003 | 50 | 0.1 | Balanced | 0.001 | 5 |
| 0.673 | 0.003 | 50 | 0.1 | Balanced | 0.001 | 10 |
| 0.812 | 0.006 | 50 | 0.05 | None | 0.001 | 5 |
| 0.813 | 0.007 | 50 | 0.05 | None | 0.001 | 10 |
| 0.673 | 0.002 | 50 | 0.05 | Balanced | 0.001 | 5 |
| 0.673 | 0.001 | 50 | 0.05 | Balanced | 0.001 | 10 |
| 0.800 | 0.004 | 50 | 0.01 | None | 0.001 | 5 |
| 0.801 | 0.002 | 50 | 0.01 | None | 0.001 | 10 |
| 0.669 | 0.002 | 50 | 0.01 | Balanced | 0.001 | 5 |
| 0.669 | 0.002 | 50 | 0.01 | Balanced | 0.001 | 10 |
| 0.829 | 0.007 | 80 | 0.15 | None | 0.001 | 5 |
| 0.827 | 0.002 | 80 | 0.15 | None | 0.001 | 10 |
| 0.676 | 0.002 | 80 | 0.15 | Balanced | 0.001 | 5 |
| 0.676 | 0.003 | 80 | 0.15 | Balanced | 0.001 | 10 |
| 0.830 | 0.004 | 80 | 0.1 | None | 0.001 | 5 |
| 0.830 | 0.005 | 80 | 0.1 | None | 0.001 | 10 |
| 0.675 | 0.005 | 80 | 0.1 | Balanced | 0.001 | 5 |
| 0.673 | 0.001 | 80 | 0.1 | Balanced | 0.001 | 10 |
| 0.826 | 0.005 | 80 | 0.05 | None | 0.001 | 5 |
| 0.829 | 0.007 | 80 | 0.05 | None | 0.001 | 10 |
| 0.672 | 0.002 | 80 | 0.05 | Balanced | 0.001 | 5 |
| 0.671 | 0.003 | 80 | 0.05 | Balanced | 0.001 | 10 |
| 0.829 | 0.004 | 100 | 0.15 | None | 0.001 | 5 |
| 0.830 | 0.002 | 100 | 0.15 | None | 0.001 | 10 |
| 0.830 | 0.002 | 100 | 0.15 | None | 0.001 | 20 |
| 0.824 | 0.016 | 100 | 0.15 | None | 0.0001 | 5 |
| 0.828 | 0.007 | 100 | 0.15 | None | 0.0001 | 10 |
| 0.830 | 0.009 | 100 | 0.15 | None | 0.0001 | 20 |
| 0.827 | 0.006 | 100 | 0.10 | None | 0.001 | 5 |
| 0.827 | 0.003 | 100 | 0.10 | None | 0.001 | 10 |
| 0.826 | 0.004 | 100 | 0.10 | None | 0.001 | 20 |
| 0.819 | 0.023 | 100 | 0.10 | None | 0.0001 | 5 |
| 0.826 | 0.008 | 100 | 0.10 | None | 0.0001 | 10 |
| 0.831 | 0.005 | 100 | 0.10 | None | 0.0001 | 20 |
| 0.821 | 0.010 | 100 | 0.05 | None | 0.001 | 5 |
| 0.823 | 0.008 | 100 | 0.05 | None | 0.001 | 10 |
| 0.823 | 0.004 | 100 | 0.05 | None | 0.001 | 20 |
| 0.813 | 0.015 | 100 | 0.05 | None | 0.0001 | 5 |
| 0.825 | 0.012 | 100 | 0.05 | None | 0.0001 | 10 |
| 0.829 | 0.006 | 100 | 0.05 | None | 0.0001 | 20 |

Tabla 5.21 Fase de desarrolo del caso de datos reducidos y modificados mediante metodos de Kernel aproximados.

En la tabla anterior se observa que los resultados mejoran mientras mayor es la cantidad de componentes utilizadas y al tener un gamma 0.1 para la aproximación de Kernel aplicada. Por lo tanto, entre los parámetros propuestos para los métodos de aproximación de Kernel, los más eficientes son gamma 0.1 y 100 componentes. Al ser los segundos mejores resultados aquellos con un gamma de 0.15. Se estima que, el verdadero valor óptimo para el gamma de Kernel debe conseguirse entre 0.1 y 0.15.

Respecto a los parámetros de SGD, hay necesidad de tener necesidad de un número máximo de iteraciones mayor para conseguir un mejor resultado. Esto se debe a que el algoritmo tiene que procesar tantas características como componentes para la aproximación de Kernel. Por lo tanto, al tener gradientes más complejas, se podría necesitar de más iteraciones para converger.

#### 5.5.3.2 Fase de evaluación

Se resalta lo siguiente, comparando con la tabla 5.15 y la 5.22:

* Globalmente existe una mejora en de 0.03 para la precisión, 0.02 para la sensibilidad y de 0.03 para el puntaje F1. Por lo tanto, se tienen más predicciones correctas.
* Para la clase de tejido no tumoral, la precisión ha mejorado en 0.02; la sensibilidad, permanece igual y el puntaje F1 ha mejorado en 0.01. Por lo tanto, se indica una reducción en la tasa de falsos positivos.
* Para la clase de tejido tumoral, la precisión ha aumentado en 0.07; la sensibilidad, en 0.12 y el puntaje F1; en 0.14. Estos valores indican una reducción los falsos positivos y falsos negativos.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Clase | Precisión | Sensibilidad | F1 | Support |
| 0 | 0.85 | 0.98 | 0.91 | 936999 |
| 1 | 0.81 | 0.33 | 0.47 | 235282 |
| Avg/tot | 0.84 | 0.85 | 0.82 | 1172281 |

Tabla 5.22 Reporte de clasificacion para el modelo de datos reducidos con aplicacion de metodos de Kernal aproximado, en la fase evaluación cuantitativa.

Comparando la matriz de confusión para el modelo seleccionado (Tabla 5.23) con la del modelo de SGD para datos reducidos (ver tabla 5.16) se puede resaltar un aumento en la cantidad total de verdaderos positivos de 85.37% el cual se refleja en la reducción de falsos negativos del 18%. Por otro lado, los verdaderos negativos y falsos positivos solo varían en un -4.09% y 0.09%, respectivamente.

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | Clase predicha | |
|  |  | No tumoral | Tumoral |
| Clase  Real | No tumoral | 918682 | 18317 |
| Tumoral | 158410 | 76872 |

Tabla 5.23 Matriz de confusion para el modelo de datos reducidos con aplicacion de metodos de Kernal aproximado, en la fase evaluación cuantitativa.

Para la **validación visual,** se visualizan las siguientes imágenes para cada uno de los cuatro pacientes seleccionados para la representación visual:

1. La clasificación real, de las muestras en el espacio de características tras aplicar el Kernel determinado en la sección anterior y una reducción de sus a dos primeras componentes principales.
2. La clasificación predicha, aplicando lo mencionado en el punto anterior.
3. La imagen de la segmentación real.
4. La imagen de la segmentación predicha.

Adicionalmente, se aporta la matriz de confusión y el reporte de clasificación para los pacientes de prueba en validación visual.

Para el **paciente 150** se observa:

* Globalmente, los niveles de precisión, sensibilidad y puntaje F1 son elevados.
* Para el tejido tumoral, se observa que sin embargo su sensibilidad es de 0.76. Entonces la cantidad de falsos negativos es un 0.24 de la cantidad de verdaderos positivos.

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | Clase predicha | |
|  |  | No tumoral | Tumoral |
| Clase  Real | No tumoral | 35823 | 35 |
| Tumoral | 316 | 1026 |

Tabla 5.24 Matriz de confusion para el paciente 150, en el caso de datos reducidos y modificados mediante métodos

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Clase | Precisión | Sensibilidad | F1 | Support |
| 0 | 0.99 | 1.00 | 1.00 | 35852 |
| 1 | 0.97 | 0.76 | 0.85 | 1342 |
| Avg/tot | 0.99 | 0.99 | 0.99 | 37200 |

Tabla 5.25 Reporte de clasificacion para el paciente 150, en el caso de datos reducidos y modificados mediante métodos de Kernel aproximados

Se puede observar en la figura 5.12 que se tiene una predicción bastante precisa. Sin embargo, se introduce un nivel de ruido que pertenece a los falsos positivos. Este ruido no se encuentra en la Figura 5.8, correspondiente al mismo paciente. El modelo no logra clasificar correctamente las muestras pertenecientes al lado inferior derecho del espacio de características de sus dos componentes principales.

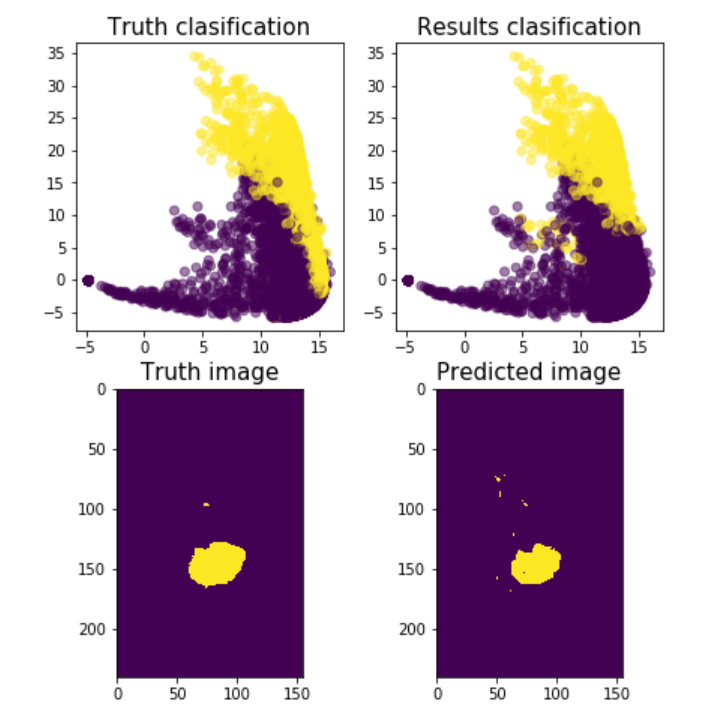


Ilustración 5.12 Validación visual del paciente 150, para el caso de datos reducidos y modificados mediante métodos de Kernel aproximados

Respecto al **paciente 185.** En la tabla 5.26 y la tabla 5.27, se determina que la sensibilidad es menor a 0.50. Por lo tanto, la cantidad de falsos negativos es mayor a la de verdaderos positivos, sin embargo se tiene que el puntaje F1 para el tejido tumoral es mayor que el caso general de la tabla 5.22.

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | Clase predicha | |
|  |  | No tumoral | Tumoral |
| Clase  Real | No tumoral | 34003 | 221 |
| Tumoral | 1719 | 1257 |

Tabla 5.26 Matriz de confusion para el paciente 185, en el caso de datos reducidos y modificados mediante métodos

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Clase | Precisión | Sensibilidad | F1 | Support |
| 0 | 0.95 | 0.99 | 0.97 | 34244 |
| 1 | 0.85 | 0.42 | 0.56 | 2976 |
| Avg/tot | 0.94 | 0.95 | 0.94 | 37200 |

Tabla 5.27 Reporte de clasificacion para el paciente 185, en el caso de datos reducidos y modificados mediante métodos

Con respecto a las imágenes, se destaca que, como en el caso de la figura 5.12 se introduce un ruido que no estaba presente en el caso de prueba anterior y este ruido aumenta la cantidad de falsos positivos. Adicionalmente, se puede ver que la proporción del tumor predicho correctamente es inferior a la mitad, sin embargo, el núcleo del tumor logra ser extraído de manera satisfactoria.

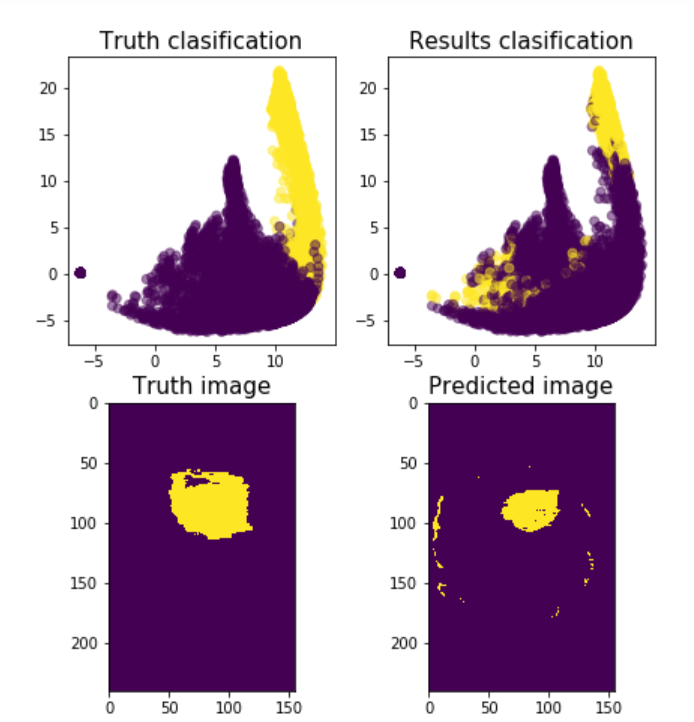


Ilustración 5.13 Validación visual del paciente 185, para el caso de datos reducidos y modificados mediante métodos de Kernel aproximados

Observando el **paciente 186,** se observa que el tejido del tumor predicho representa un 66% del tejido tumoral total. Por lo tanto, a diferencia del paciente anterior, la segmentación es más efectiva. En la validación visual de la figura 5.14, se repite el mismo fenómeno donde se agrega un ruido que no estaba presente en la etapa anterior.

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | Clase predicha | |
|  |  | No tumoral | Tumoral |
| Clase  Real | No tumoral | 34675 | 95 |
| Tumoral | 829 | 1601 |

Tabla 5.28 Matriz de confusion para el paciente 186, en el caso de datos reducidos y modificados mediante métodos

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Clase | Precisión | Sensibilidad | F1 | Support |
| 0 | 0.98 | 1.00 | 0.99 | 34770 |
| 1 | 0.94 | 0.66 | 0.78 | 2430 |
| Avg/tot | 0.97 | 0.98 | 0.97 | 37200 |

Tabla 5.29 Reporte de clasificacion para el paciente 186, en el caso de datos reducidos y modificados mediante métodos

.

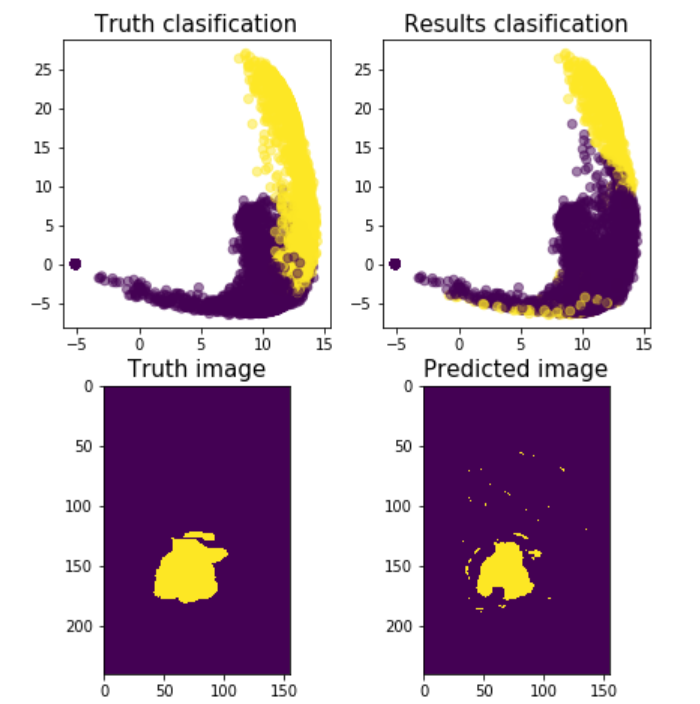


Ilustración 5.14 Validación visual del paciente 186, para el caso de datos reducidos y modificados mediante métodos de Kernel aproximados

Finalmente, el **paciente 205** refleja un puntaje una segmentación donde solo el 46% del volumen tumoral es identificado. Sin embargo, cabe destacar que en ninguno de los tres casos de prueba se logró segmentar este caso específico con una segmentación eficiente. Como se puede observar en la imagen 5.15 el mismo fenómeno de ruido se repite para dicho caso.

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | Clase predicha | |
|  |  | No tumoral | Tumoral |
| Clase  Real | No tumoral | 34608 | 53 |
| Tumoral | 1360 | 1179 |

Tabla 5.30 Matriz de confusion para el paciente 205, en el caso de datos reducidos y modificados mediante métodos

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Clase | Precisión | Sensibilidad | F1 | Support |
| 0 | 0.96 | 1.00 | 0.98 | 34661 |
| 1 | 0.96 | 0.46 | 0.63 | 2539 |
| Avg/tot | 0.96 | 0.96 | 0.96 | 37200 |

Tabla 5.31 Reporte de clasificacion para el paciente 205, en el caso de datos reducidos y modificados mediante métodos

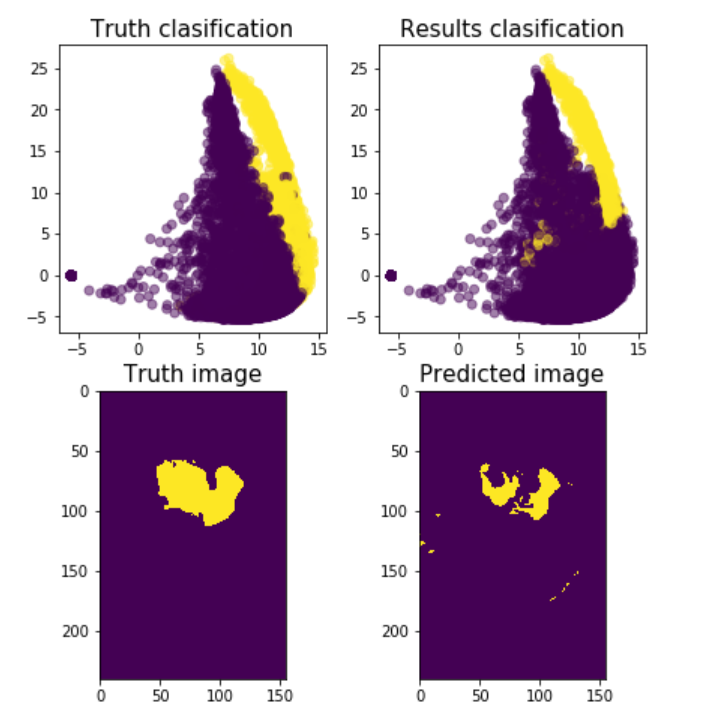


Ilustración 5.15 Validacion del paciente 205, para el caso de datos reducidos y modificados mediante metodos de Kernel aproximados.

### 5.5.4 Recapitulación

Para recapitular, todas las comparaciones y análisis que se establecieron a lo largo capitulo, el mejor modelo establecido es aquel que utiliza como entrada aquel que, tras la aplicación del esquema de reducción expuesto en el **capítulo 4,** se aplican los métodos aproximados de Kernel y el análisis de componentes principales. Los hiperparametros establecidos como óptimos se hallan en la siguiente tabla.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Hiperparametros de Kernel | | Hiperparametros de SGD | | |
| Número de componentes | Gamma | Balance entre clases | Alpha | Número máximo de iteraciones |
| 100 | 0.1 | No | 0.0001 | 20 |

Tabla 5.32 Hiperparametros seleccionados como óptimos

Los resultados de estos parámetros ya han sido analizados a lo largo del **capítulo 5.5.3**. En líneas generales, si se observan los resultados cuantitativos aun es necesario mejorar la sensibilidad del algoritmo para obtener mejores puntajes F1. Por otro lado, observando las imágenes de las segmentaciones, se observa que el borde del tumor no es predicho correctamente como se hace con el centro del tumor. Normalmente, el centro del tumor está más relacionado a los tejidos de tumor necrótico y tumor realzado. Mientras que la parte exterior está más relacionada a los edemas. En la **sección 4.2.1** ya se había previsto que existiría una dificultad en analizar este tejido. Sin embargo, se necesita un análisis más profundo donde se consideren todas las clases originales del dataset para determinar cuál es el tejido más difícil de predecir.

A pesar de que este modelo aporte ruido a la predicción, este puede ser fácilmente corregido mediante la implementación de algoritmos básicos de procesamiento de imágenes o mediante algoritmos de agrupamiento.

# 

Conclusiones y Recomendaciones

En los capítulos anteriores, se expusieron los principios y métodos básicos de la segmentación de tumores cerebrales en imágenes de resonancia magnética. Se analizaron los diferentes métodos según su clasificación y sus principales ventajas y desventajas. Con el fin de entender qué modelo se puede aplicar, de acuerdo a los recursos de hardware que se tienen para ejecutar este proyecto (Ver Apéndice A).

Enseguida se realizó un recuento de cuáles son los conceptos teóricos con los que se trabajó a lo largo del proyecto. Estos tuvieron un enfoque matemático que permitieron establecer las bases sobre el tema y realizar una presentación correcta del sistema. Entonces, se permitió a realizar un estudio estadístico de los datos correspondiente a los encéfalos de 210 pacientes de tumores de alto grado, con el fin de asegurar la factibilidad de establecer los métodos de clasificación de pixeles para alcanzar una predicción aceptable, además dicho estudio estadístico permitió prever qué tipo de tejidos presentarían mayores dificultades para predecir que, en este caso fue, el edema. Dicho estudio pre-exploratorio de los datos también se acompañó de un estudio cualitativo, con el fin de visualizar la distribución de los datos. A pesar de que solo se pudieron utilizar tres de las cuatro características disponibles, debido a las limitaciones del ojo humano.

Adicionalmente, se analizó que el volumen de los datos puede traer como consecuencias tiempos de entrenamiento muy largos, y por lo tanto, se confirmó la necesidad de utilizar métodos estocásticos para permitir un análisis aceptable y con resultados similares a sus homólogos. Otro estudio pre-exploratorio realizado, permitió determinar la proporción de los tejidos tumorales respecto a los tejidos sanos y, entonces, se definió como prioridad reducir el desbalance entre ambas clases. Gracias a este análisis pre-exploratorio se pudo justificar gran parte de los métodos utilizados; por lo tanto, esto ha permitido alcanzar su objetivo específico y su ejecución ha sido clave para el resto del trabajo.

Para actuar ante parte de las problemáticas expuestas en dicho análisis pre-exploratorio, se seleccionaron solamente los cortes óptimos de los volúmenes en 3D y, en seguida, se decidió utilizar el método de reducción de datos establecido por Shen [37], que consta de una reducción de clústeres redundantes y una reducción de los datos redundantes dentro de los clústeres restantes. Sin embargo, tras algunas ejecuciones se observó que el algoritmo no se adapta a la gran variabilidad de las diferentes clases dentro del espacio de características. Por lo tanto, se establecieron modificaciones que permitan corregir los defectos de dicho algoritmo con el dataset y, de esta manera, se logró una implementación más eficiente en términos de reducción de números de datos y de equilibrio entre las clases. La reducción de datos permite reducir hasta en un 83.7 % el tiempo de aprendizaje del algoritmo de aprendizaje estocástico. Por lo tanto, tras el establecimiento de dichos algoritmos de reducción se puede considerar que el objetivo específico correspondiente ha sido cumplido, ya que este algoritmo sirve a un pre procesamiento que permite mejorar los datos para la ejecución del algoritmo de aprendizaje.

A continuación, se presentó el algoritmo de aprendizaje ejecutado y más técnicas de pre-procesamiento de datos como los métodos aproximados de Kernel y la aplicación del análisis de componentes principales. Se estableció un esquema de validación cruzada que permite hallar los hiperparametros con mejor desempeño para algoritmo de SGD y, así, reducir el sobreajuste. Adicionalmente, se establecieron tres métricas diferentes a optimizar para la validación cruzada. La validación cruzada, permite analizar de una manera preliminar el rendimiento del algoritmo con una pequeña porción de datos no conocida. Por lo tanto, con los valores hallados se observa que el algoritmo es capaz de predecir, al menos, una porción del tejido cerebral.

Al momento de poner a prueba el algoritmo de aprendizaje, con el fin de comparar el desempeño de la reducción de datos, se ejecutó un caso de prueba donde se omite el pre-procesamiento de los datos y otro donde se ejecuta. Se concluyó que la reducción de datos es más eficiente, siempre que se decida optimizar la validación cruzada en cuanto en la precisión. Por lo tanto, los algoritmos de reducción aplicados no solo fueron más eficientes en cuanto a tiempo, balance de muestras ni volumen de datos, sino en desempeño.

El último caso de prueba introdujo la aplicación de los métodos de Kernel aproximados que permitieron mejorar las métricas de clasificación de los casos de Kernel lineal. Para la fase de evaluación de los diferentes algoritmos planteados, se permitió tener una evaluación cuantitativa mediante el uso de matrices de confusión y de reportes de clasificación. Al mismo tiempo, se propuso una validación visual mediante las imágenes de pacientes aleatorios. Este contraste entre la evaluación cuantitativa y la evaluación cualitativa, permitió confirmar que mejores puntajes no siempre significan mejores resultados en cuanto a la segmentación del tumor cerebral.

Por ejemplo, la validación visual dio a conocer que el entrenamiento a partir de los datos reducidos fracasa si se tiene como objetivo optimizar la sensibilidad o el puntaje F1, a pesar de que sus puntajes de segmentación sean superiores que en el caso de que se desee estabilizar la precisión. Este análisis de resultados permitió observar los resultados desde un punto de vista ambivalente y se considera que cumplió con los objetivos pautados.

Sin embargo, con el fin de mejorar los resultados de este trabajo o de enriquecer lo que ha sido implementado se proponen las siguientes alternativas:

* Extender el perímetro de las clases a clasificar, incluyendo a la clasificación las clases de tumor necrótico, tumor realzado y de edema. Con el fin de analizar más detalladamente en cuales de esos tejidos el algoritmo de segmentación automática puede representar fallas.
* El espacio de características de este estudio ha sido restringido a cuatro. Sin embargo, es posible agregar más características a dicho espacio, como por ejemplo, las gradientes de diferentes pixeles o voxeles con respecto a sus vecinos. El uso de las gradientes permite también agregar datos que implican la espacialidad de los pixeles o voxeles, en lugar de basarse solamente en términos de intensidades.
* Otra posibilidad, es la de intentar, con otro algoritmo de clasificación de pixeles, el mismo modelo de reducción de datos y observar si este también puede aportar mejoras en la eficiencia de dicho algoritmo en caso de no ser reducido.
* Se puede agregar otra etapa suplementaria que use como entrada las predicciones realizadas por dichos algoritmos de aprendizaje. Esta nueva etapa puede conseguir de la utilización de técnicas básicas de procesamiento de imágenes que permitan corregir las predicciones previamente realizadas. Por ejemplo, hallar los bordes de los tumores mediante análisis de gradientes y tratar de hacer coincidir la plantilla que se obtuvo a partir de las diferentes predicciones.
* Con un GPU que pueda procesar una mayor cantidad de datos con tiempos de entrenamiento más aceptables, se puede extender las fases de entrenamiento, validación y evaluación a volúmenes completes en 3D de los pacientes, para observar el desempeño de los resultados más allá de los cortes óptimos.

Con el fin de fomentar la continuación de este proyecto, el código comentado y los datasets obtenidos en las diferentes etapas del proyecto se pueden conseguir en el repositorio que se encuentra en **el apéndice C**.

# 

Referencias

1. Nelly Gordillo, Eduard Montseny, Pilar Sobrevilla, State of the art survey on MRI brain tumor segmentation, Magnetic Resonance Imaging, Volumen 31, Numero 8, 2013, Paginas 1426-1438, ISSN 0730-725X.
2. Mazzara G, Velthuizen R, Pearlman J, Greenberg H, Wagner H. Brain tumor target determination or radiation treatment planning throught automated MRI segmentation. Int J Radiat Oncol Biol Phys 2004; 59(1):300-12.
3. Luo S, Li R, Ourselin S. A new deformable model using dynamic gradient vector flow and adaptive balloon forces. APRS Workshop on Dig Image Comp; 2003.p. 9–14.
4. Spyridon Bakas et al, Identifying the Best Machine Learning Algorithms for Brain Tumor Segmentation, Progression Assessment, and Overall Survival Prediction in the BRATS Challenge
5. S. Theodoridis, K. Koutroumbas, Pattern Recognition(4nd ed.), Academic Press, New York, 2009.
6. Dennis and Schnabel , “Numerical methods for unconstrained optimization and nonlinear equations”, 1983
7. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby JS, Freymann JB, Farahani K, Davatzikos C. "Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features", Nature Scientific Data, 4:170117 (2017) DOI: 10.1038/sdata.2017.117
8. Menze BH, Jakab A, Bauer S, Kalpathy-Cramer J, Farahani K, Kirby J, Burren Y, Porz N, Slotboom J, Wiest R, Lanczi L, Gerstner E, Weber MA, Arbel T, Avants BB, Ayache N, Buendia P, Collins DL, Cordier N, Corso JJ, Criminisi A, Das T, Delingette H, Demiralp Γ, Durst CR, Dojat M, Doyle S, Festa J, Forbes F, Geremia E, Glocker B, Golland P, Guo X, Hamamci A, Iftekharuddin KM, Jena R, John NM, Konukoglu E, Lashkari D, Mariz JA, Meier R, Pereira S, Precup D, Price SJ, Raviv TR, Reza SM, Ryan M, Sarikaya D, Schwartz L, Shin HC, Shotton J, Silva CA, Sousa N, Subbanna NK, Szekely G, Taylor TJ, Thomas OM, Tustison NJ, Unal G, Vasseur F, Wintermark M, Ye DH, Zhao L, Zhao B, Zikic D, Prastawa M, Reyes M, Van Leemput K. "The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)", IEEE Transactions on Medical Imaging 34(10), 1993-2024 (2015) DOI: 10.1109/TMI.2014.2377694
9. Definicion de "[overfitting](https://en.oxforddictionaries.com/definition/overfitting)" at [OxfordDictionaries.com](https://en.wikipedia.org/wiki/OxfordDictionaries.com): Esta definicion es especial para estadistica.
10. Wu, Y *automatic Segmentation of Brain Tumors Using Population and Individual Information. Journal of Digital*., Yang, W., Jiang, J., Li, S., Feng, Q., & Chen, W. (2013). *Semi- Imaging, 26(4), 786–796.*doi:10.1007/s10278-012-9568-1
11. Foo JL. A survey of user interaction and automation in medical image segmentation methods. Tech rep ISUHCI20062, Human Computer Interaction Department, Iowa State Univ; 2006.
12. Schmidt M. Automatic brain tumor segmentation. Master’s thesis, University of Alberta, 2005
13. Wang, G., Li, W., Ourselin, S., & Vercauteren, T. (2019). *Automatic Brain Tumor Segmentation Using Convolutional Neural Networks with Test-Time Augmentation. Lecture Notes in Computer Science, 61–72.*doi:10.1007/978-3-030-11726-9\_6
14. Navpreet kaur, Mr. Manvinder Sharma, Brain Tumor Segmentation Techniques: A Survey, International Journal of Modern Computer Science (IJMCS), Volume 4, Issue 4, August, 2016, p (220 – 223)
15. Sonia Pujol, Automatic Segmentation of Brain Structures, Surgical Planning Laboratory, Harvard Medical School. Disponible en: <https://www.slicer.org/w/images/2/24/AutomaticSegmentation_SoniaPujol.pdf>
16. R. Fisher, S. Perkins, A. Walker and E. Wolfart. *Adaptative Thresholding*. 2003. Disponible en: <http://homepages.inf.ed.ac.uk/rbf/HIPR2/adpthrsh.htm>
17. S. Theodoridis, K. Koutroumbas, Pattern Recognition(4nd ed.), Academic Press, New York, 2009.
18. Devijver, P. A., and J. Kittler, *Pattern Recognition: A Statistical Approach, Prentice-Hall*, Londres, 1982
19. Harish Kandan. (2017, Agosto) Towards Data Science. [Online]. <https://towardsdatascience.com/understanding-the-kernel-trick-e0bc6112ef78>
20. Seiichi Uchida, *Image processing and recognition for biological images*, 2013; Disponible en: <https://www.researchgate.net/publication/236125496_Image_processing_and_recognition_for_biological_images>
21. Manjot Kaur , Pratibha Goyal, *A Review on Region Based Segmentation,* International Journal of Science and Research (IJSR), ISSN (Online): 2319-7064, (2013)
22. ["Magnetic Resonance Imaging"](https://web.archive.org/web/20170510065614/https:/www.radiology.wisc.edu/education/med_students/neuroradiology/NeuroRad/Intro/MRIintro.htm). [University of Wisconsin](https://en.wikipedia.org/wiki/University_of_Wisconsin). Archived from [the original](https://www.radiology.wisc.edu/education/med_students/neuroradiology/NeuroRad/Intro/MRIintro.htm) on 2017-05-10. Retrieved 2016-03-14.
23. J. Cousty, *MorphoGraph and Imagery,* 2011. Disponible en <https://perso.esiee.fr/~coustyj/EnglishMorphoGraph/L10.pdf>
24. Anton S. Kornilov, Ilia V. Safonov, *An Overview of Watershed Algorithm*
25. *Implementations in Open Source Libraries*, 2018 Disponible en :

[www.esiee.fr/~coustyj/EnglishMorphoGraph/L10.pdf](http://www.esiee.fr/~coustyj/EnglishMorphoGraph/L10.pdf)

1. K. Kamnitsas, W. Bai , E. Ferrante , S. McDonagh , M. Sinclair, N. Pawlowski, M. Rajchl, M. Lee, B. Kainz, D. Rueckert, B. Glocke, *Ensembles of Multiple Models and Architectures for Robust Brain Tumour Segmentation*, Biomedical Image Analysis Group, Imperial College London (2017) Disponible en: <https://arxiv.org/pdf/1711.01468.pdf>
2. M., Bishop, Christopher (1995). *Neural networks for pattern recognition*. Clarendon Press. [ISBN](https://en.wikipedia.org/wiki/International_Standard_Book_Number) [978-0198538493](https://en.wikipedia.org/wiki/Special:BookSources/978-0198538493).
3. Tim McInerney and Demetri Terzopoulos, *Deformable Models in Medical Image Analysis: A Survey,* 2000
4. Instituto Nacional de Cancer de los Estados Unidos, Grados de un tumor, (2013) Disponible en: <https://www.cancer.gov/espanol/cancer/diagnostico-estadificacion/pronostico/hoja-informativa-grado-tumor>
5. American Joint Committee on Cancer. AJCC Cancer Staging Manual. 7th ed. New York, NY: Springer; 2010.
6. David N. Louis, Arie Perry, Guido Reifenberger, Andreas von Deimling, Dominique Figarella‑Branger, Webster K. Cavenee, Hiroko Ohgaki, Otmar D. Wiestler, Paul Kleihues, David W. Ellison, *The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary,* (2016) Disponible en: <https://braintumor.org/wp-content/assets/WHO-Central-Nervous-System-Tumor-Classification.pdf>
7. Joseph V. Hajnal et al, *Use of Fluid Attenuated Inversion Recovery (FLAIR) Pulse Sequences in MRI of the Brain, (1992)* Journal of Computer Assisted Tomohraphy,. Disponible en : <https://mriquestions.com/uploads/3/4/5/7/34572113/use_of_fluid_attenuated_inversion_recovery__flair_.1.pdf>
8. Meghana M Chavan, Asawari Patil, Lata Dalvi, Ajinkya Patil, *Mini Batch K-Means Clustering On Large Dataset,* International Journal of Scientific Engineering and Technology Research, Vol. 4, Issue.07, (2015) ( p1356-1358). Disponible en: <http://ijsetr.com/uploads/462153IJSETR4282-245.pdf>
9. Aurea Grané, *Analisis de Componentes Principales*, Universidad Carlos III de Madrid. Disponible en: <http://halweb.uc3m.es/esp/Personal/personas/agrane/ficheros_docencia/MULTIVARIANT/slides_comp_reducido.pdf>
10. Bob Hayes, Programming Languages Most Used and Recommended by Data Scientists, (2019) Disponible en: <https://businessoverbroadway.com/2019/01/13/programming-languages-most-used-and-recommended-by-data-scientists/>
11. Sitio official de numPy: <https://www.numpy.org/>
12. Denis Cousineau, Sylvain Chartier, Outliers detection and treatment : a review. International Journal of Physchological Research, 2010, Vol 3 No. 1 (p 58-67) Disponible en: <https://dialnet.unirioja.es/descarga/articulo/3296435.pdf>
13. Xiang-Jun Shen, Lei Mu, Zhen Li, Hao-Xiang Wu, Jian-Ping Gou, Xin Chen, *Large-scale support vector machine classification with redundant data reduction*, Neurocomputing, Volumen 172, 8 Enero 2016, P 189-197
14. Luis Moreno Sanchez, Glioblastoma: Multidisciplinary treatment approaches, 2017, <https://www.researchgate.net/figure/Progression-in-1-month-of-untreated-GBM_fig1_318598475>
15. Mohtadi Ben Fraj, In Depth: Parameter tuning for SVC, 2018, Disponible en: <https://medium.com/all-things-ai/in-depth-parameter-tuning-for-svc-758215394769>
16. Hugo Salinas, *Clase 2: Estadistica*, Universidad de Atacama – Chile, (2011), Disponible en: <http://www.mat.uda.cl/hsalinas/cursos/2011/2do/clase2.pdf>
17. Lawrence T. DeCarlo, On the Meaning and Use of Kurtosis, Physchological Methods, 1997, Vol. 2, No. 3, (p. 292 – 307) Disponible en: <https://pdfs.semanticscholar.org/6f7e/f8ed60553b37bdf335876ac4e0adf400d1a8.pdf>
18. Confusion matrix (en ingles), Wikipedia. Disponible en : <https://en.wikipedia.org/wiki/Confusion_matrix>
19. Tom Fawcett, an introduction to ROC analysis, Pattern Recognition Letters 27 (2006) 861–874, disponible en: <https://people.inf.elte.hu/kiss/11dwhdm/roc.pdf>
20. David M W Powers, Evaluation: From Precision, Recall and F-Factor to ROC, Informedness, Markedness & Correlation, Technical Report SIE-07-001, (2007) Disponible en: <http://www.flinders.edu.au/science_engineering/fms/School-CSEM/publications/tech_reps-research_artfcts/TRRA_2007.pdf>
21. Documentacion de sci-kit learn. *Classification Report.* Disponible en : <https://scikit-learn.org/stable/modules/generated/sklearn.metrics.classification_report.html>
22. Documentacion de sci-kit learn, Approximative Kernels, Disponible en : <https://scikit-learn.org/stable/modules/kernel_approximation.html>
23. Rahimi, A. and Recht, B, “Random features for large-scale kernel machines”, Advances in neural information processing, (2007), Disponible en : <https://people.eecs.berkeley.edu/~brecht/papers/07.rah.rec.nips.pdf>
24. Documentacion de sci-kit learn, *Grid Search*, Disponible en: <https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html>
25. Rohan Joseph, *Grid Search for model Tuning,* Towards data science, (2018), Disponible en: <https://towardsdatascience.com/grid-search-for-model-tuning-3319b259367e>

## Apéndice A

## Python



Pyhton es un lenguaje de programación interpretado inventado en 1991. Es un lenguaje de programación multiparadigma debido a que permite al programador utilizar diferentes estilos de programación como:

* Programación orientada a objetos
* Programación imperativa
* Programación funcional
* Otros (Mediante el uso de extensiones)

Adicionalmente, es el lenguaje de programación más utilizado y, al mismo tiempo, el más recomendado para la ciencia de datos y aprendizaje artificial; Según una encuesta realizada por la comunidad de científicos de datos de *Kaggle* en 2019.

### Numpy

NumPy es una extensión de Python, que permite tener un mejor soporte para vectores y matrices de gran tamaño y junto a una gran colección de funciones matemáticas de alto nivel.

### Pandas

Pandas es una librería *open-source y extensión de NumPy* que permite utilizar estructuras de datos fácil de usar y que son compatible con la mayoría de las herramientas de análisis de datos de Python. Las estructuras de datos disponibles en Pandas permiten trabajar de una forma más sencilla que con los típicos diccionarios o arreglos que son utilizados por defecto en Python. Su nombre proviene del término de econometría, “Datos de Panel”.

### Sci-kit Learn

### Matplob lib

## Herramientas de hardware

Durante el desarrollo del código se utilizó un computador con las siguientes características durante la fase de desarrollo y de evaluación:

* **Procesador:** Intel® Core™ i5-8350U CPU @ 1.70 GHz – 1.90 GHz.
* **Memoria RAM**: 8 GB (7.88 GB utilizables)
* S**istema de explotación**: 64 bits

## Apéndice B

### Análisis pre-exploratorio detallado de cada paciente

A continuación, se presenta el documento Excel donde se extrajeron los datos de los pacientes en 3D. Se presentan las siguientes pestanas dentro del documento:

* **Resumen**: Donde se exponen las mismas tablas que en la **sección 4.2.1** y la **sección 4.2.4.**
* **T1\_Stats, T1C\_Stats, T2\_Stats, FLAIR\_Stats :** En estas pestanas se expone individualmente por paciente y segun la modalidad correspondiente, los siguientes datos:
  + Media
  + Varianza
  + Mediana
  + Asimetria estadística
  + Curtosis
  + Cantidad de voxeles correspondiente a dicha clase (Tejido tumoral, tejido sano, edema, tumor necrótico o tumor realzado)



## Apéndice C

### Código comentado

El código comentado que se ha utilizado a lo largo del proyecto, se puede conseguir en el siguiente enlace:

<https://github.com/JuanOcando/SegmentacionIRM>