

UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

POSGRADO EN CIENCIAS FÍSICAS

MÉTODOS DE APRENDIZAJE SUPERVISADO EN IMÁGENES PET DE
PACIENTE CON ENFERMEDAD DE ALZHEIMER



TESIS QUE PARA OPTAR POR EL GRADO DE:

MAESTRO EN CIENCIAS (FÍSICA MÉDICA)

PRESENTA:

**I.F. MOISÉS EBENEZER HERNÁNDEZ
CRUZ**

COMITÉ TUTOR

**DR. ARTURO AVENDAÑO ESTRADA
FACULTAD DE MEDICINA**

**DR. ARNULFO MARTÍNEZ DÁVALOS
INSTITUTO DE FÍSICA**

**DR. MIGUEL ÁNGEL ÁVILA RODRÍGUEZ
FACULTAD DE MEDICINA**

**DRA. LIZBETH ROSSANA AYALA DOMÍNGUEZ
UNIVERSIDAD DE WISCONSIN-MADISON**

CIUDAD DE MÉXICO,

JUNIO DE 2023



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

*Para mi abuelita, Tita.
Para mi amor, Geena.*

Agradecimientos institucionales

Agradezco al Consejo Nacional de Humanidades Ciencias y Tecnologías (CONAHCYT) por la beca otorgada para cursar la maestría en ciencias (física médica) y poder realizar este proyecto.

Al Posgrado en Ciencias Físicas de la UNAM por el apoyo brindado durante la maestría y la realización de este proyecto, así como al personal académico y administrativo.

Al Programa de Apoyo a los Estudios de Posgrado (PAEP) por los recursos brindados para presentar un borrador de este proyecto durante el LXV Congreso Nacional de Física celebrado en el mes de octubre de 2022 en la ciudad de Zacatecas, Zacatecas, México.

A los profesores de la maestría en ciencias (física médica) por los conocimientos impartidos.

Al personal de la Unidad PET/CT de la Facultad de Medicina de la UNAM por permitir el acceso a su base de datos y por el apoyo brindado durante la realización de este proyecto.

A la Unidad de Radiofarmacia-Ciclotrón de la UNAM por permitir el uso de sus instalaciones durante el desarrollo de este proyecto, así como al personal del Laboratorio de Investigación Preclínica microPET por facilitar el uso de los equipos y software.

A los miembros de mi comité tutor por el apoyo y orientación brindados durante la realización de este trabajo: Dr. Arturo Avendaño Estrada, Dr. Arnulfo Martínez Dávalos, Dr. Miguel Ángel Ávila Rodríguez y Dra. Lizbeth Rossana Ayala Domínguez.

A los sinodales que amablemente revisaron este trabajo para su aprobación: Dr. Arturo Avendaño Estrada, Dra. María Julieta Mateos Ochoa, Dra. Janeth Serrano Bello, Dra. Ruth Fuentes García y Dr. Jaime Fabian Vázquez de la Rosa.

Agradecimientos personales

La finalización de este trabajo no sería posible sin el apoyo que recibí de varias personas, a quienes les agradezco mucho, en especial a quienes me acompañaron a lo largo de este trayecto.

A mi tutor, el Dr. Arturo Avendaño Estrada, le agradezco por ayudarme a obtener las bases de conocimiento necesarias, por guiarme y apoyarme a lo largo de la maestría y por motivarme siempre a terminar este trabajo en tiempo y forma.

A mis padres, Víctor y Jesie, les agradezco por brindarme los medios para mi formación básica y sobretodo por su amor y cariño, que siempre llevo en mi corazón. También agradezco a mi hermano Habid, a mi tía Mariela, a mi tía Keren, a mi familia y familia de mi novia por apoyarme de diferentes formas en estos años, son gestos que siempre tendré presentes.

A la señora Verónica y a mis amigos Alexander y Natalia por todo el apoyo y cariño que me brindaron, hicieron que mi estancia en un lugar nuevo sea mucho más fácil y estoy agradecido por ello.

A mis compañeros y amigos de la maestría: Jaz, Fer, Miriam, Santi, Ana, Mau, Irene, Tony y Erick. Les agradezco por todas las veces que me ayudaron durante la maestría, incluso fuera de lo académico, en especial a quienes desde mi llegada me apoyaron y motivaron, siempre les estaré agradecido por ello y por los lazos de amistad forjados.

Agradezco a mi tía Arsenia por apoyarme sin pedir nada a cambio desde mi formación media hasta hoy que finalizo esta etapa en el posgrado, así como también a mi bisabuela Olga por ayudar a que esto sea posible, siempre llevaré conmigo esas acciones y las enseñanzas y cariño de ambas.

A Geena por ser junto con mi abuelita mi mayor apoyo en todos los aspectos durante esta etapa, porque desde que decidí realizar esta maestría siempre me animó a dar mi mayor esfuerzo a pesar de las adversidades y por ayudarme siempre que lo necesité, aún estando lejos. Agradezco por todo el amor y paciencia, que espero brindarte de vuelta por mucho tiempo. Por último pero no

menos importante, a mi abuelita, Tita, por todo el apoyo brindado no solo en esta etapa sino desde que tengo memoria, lo cual ha sido fundamental para que el día de hoy pueda ser profesionista, no me alcanzará la vida para agradecerte a ti y a Dios porque has sido como un ángel para mi hermano y para mi, pero espero retribuir un poco dedicándote este trabajo.

”Esfuézate y sé valiente, no temas ni desmayes, pues yo soy el señor tu Dios y estaré contigo por donde quiera que vayas”.

Índice de abreviaturas

A β : Amiloide beta.

Alz: Pacientes con sospecha de Enfermedad de Alzheimer.

AUC: (Area Under the Curve) Área bajo la curva.

COF: Corte orbitofrontal.

CT: Tomografía Computarizada.

DNN: (Deep Neural Network) Red neuronal profunda.

DSM-IV: Diagnostic and Statistical Manual of Mental Disorders.

EA: Enfermedad de Alzheimer.

EC: Exactitud de clasificación.

EnfNeu: Pacientes con alguna enfermedad neurológica.

FDG: Fluorodesoxiglucosa.

FOV: (Field of View) Campo de Vista.

GCA: Giro del cíngulo anterior.

GF: Giro frontal.

GFI: Giro frontal inferior.

GFM: Giro frontal medial.

GFS: Giro frontal superior.

GL: Giro lingual.

GLCM: (Grey Level Co-occurrence Matrix) Matriz de co-ocurrencia de nivel de gris.

GLDZM: (Grey Level Distance Zone Matrix) Matriz de zona de distancia de nivel de gris.

GLRLM: (Grey Level Run Length Matrix) Matriz de longitud de carrera de nivel de gris.

GLSZM: (Grey Level Size Zone Matrix) Matriz de zona de tamaño de nivel de gris.

GLZLM: (Grey Level Zone Length Matrix) Matriz de longitud de la región de nivel gris.

GOP: Giro orbital posterior.

GPS: Giro parietal superior.

GR: Giro recto.

IBSI: (Image Biomarker Standardisation Initiative) Iniciativa de estandarización de biomarcadores de imagen.

KNN: (K-Nearest Neighbor) K vecino más cercano.

LF: Lóbulo frontal.

LO: Lóbulo occipital.

LOL: Lóbulo occipital lateral.

LOR: (Line Of Response) Línea de Respuesta.

LP: Lóbulo parietal.

LRHGLE: (Long Run High Grey Level Emphasis) Énfasis en niveles de gris altos de corrida larga.

LT: Lóbulo temporal.

LTA: Lóbulo temporal anterior.

LZHGLE: (Large Zone High Grey Level Emphasis) Énfasis en niveles de gris altos en zonas grandes.

MCI: (Mild Cognitive Impairment) Deterioro Cognitivo Leve.

MLP: (Multi-Layer Perceptron) Perceptrón multi capa.

MMSE: Mini Mental Status Examination.

NGLDM: (Neighbouring Grey Level Dependence Matrix) Matriz de dependencia del nivel de gris vecino.

NGTDM: (Neighbourhood Grey Tone difference Matrix) Matriz de diferencia de tono de gris del vecindario.

NID: (Normalised Inverse Difference) Diferencia inversa normalizada.

PC: Pacientes control.

PET: (Positron Emission Tomography) Tomografía por Emisión de Positrones.

ROC: (Receiver Operating Characteristic) Característica operativa del receptor.

ROI: (Region of Interest) Región de Interés.

SRTM: (Simplified Reference Tissue Model) Modelo de Tejido de Referencia Simplificado.

SUV: (Standarized Uptake Value) Valor de Captación Estandarizado.

SVM: (Support Vector Machine) Máquinas de soporte vectorial.

TOF: (Time of Flight) Tiempo de vuelo.

VOI: (Volume of Interest) Volumen de Interés.

Resumen

Las demencias son un grupo de enfermedades que afectan principalmente a la población senil y que, de manera general, ocasionan alteración de las capacidades cognitivas e intelectuales. Una de las demencias que se presentan con mayor frecuencia, principalmente después de los 60 años, es la enfermedad de Alzheimer (EA), la cual es una enfermedad neurodegenerativa progresiva que se caracteriza por una pérdida de funciones cognitivas y memoria. Como parte de la fisiopatología de la enfermedad, uno de los cambios metabólicos importantes observados en el cerebro de pacientes con Alzheimer es una disminución en el metabolismo de glucosa. Dicho cambio es consecuencia de lesiones en estructuras cerebrales ocasionadas por la acumulación de β amiloide, un péptido que forma placas seniles, y de la proteína tau. Estas lesiones afectan la regulación, absorción y utilización de glucosa. Las técnicas de imagen médica se pueden utilizar como apoyo en el diagnóstico de la EA para valorar los cambios anatómicos y fisiológicos en los cerebros de los pacientes, siendo la Tomografía por Emisión de Positrones (PET, por sus siglas en inglés), una técnica que permite evaluar cambios metabólicos. El radiodiagnóstico tiende cada vez más a apoyarse en herramientas computacionales de análisis de imagen que permiten asistir a los médicos en la toma de decisiones. Una de estas técnicas es la radiómica, la cual consiste en obtener matrices de datos a partir de la información de intensidad y forma en una región de interés. Las características radiómicas han demostrado su utilidad para obtener información importante en muchas enfermedades, por ejemplo, en enfermedades neurológicas la radiómica se ha utilizado para medir características en regiones cerebrales que tienen implicaciones con alguna enfermedad específica. De los análisis con radiómica, se pueden obtener grandes cantidades de datos que deben ser analizados por métodos estadísticos y nuevas técnicas de ciencia de datos. El análisis y ciencia de datos es una de las áreas con mayor importancia en la actualidad ya que en muchos sectores, incluyendo las ciencias de la salud, se utilizan grandes cantidades de datos que deben ser gestionados adecuadamente. Esta necesidad ha impulsado el uso de herramientas de aprendizaje autónomo que permitan identificar patrones en los datos, realizar estimaciones e incluso en algunos casos realizar la toma de decisiones. En el estudio de la EA, hay algunos trabajos acerca del uso de características radiómicas medidas en imágenes PET para entrenar algoritmos de aprendizaje autónomo y clasificar pacientes con y sin la enfermedad, sin embargo, se requieren más estudios utilizando diferentes metodologías para medir las características radiómicas y entrenar o ajustar un modelo que funcione adecuadamente como apoyo en el diagnóstico de la EA.

En este trabajo se utilizó una muestra de 75 estudios [^{18}F]FDG-PET: 25 pacientes con sospecha de EA, 25 pacientes con alguna enfermedad neurológica (EN) y 25 sujetos control sin enfermedad neurológica. Todos los estudios fueron adquiridos retrospectivamente de la Unidad PET/CT de la UNAM. Se realizaron dos estudios considerando el diagnóstico de envío de los pacientes como clase objetivo: utilizando pacientes con sospecha de EA y pacientes con alguna

EN, y utilizando pacientes con sospecha de EA y sujeto control. Además, se realizó un tercer estudio utilizando de nueva cuenta pacientes con sospecha de EA y controles, pero considerando como clase objetivo la clasificación dada por una herramienta de análisis de Alzheimer de un software comercial. De esta manera, en cada análisis se tuvieron 50 pacientes. Todas las imágenes PET se normalizaron al espacio estandar usando un atlas cerebral (HAMERS-N30R83) y luego se midieron las características radiómicas en cada volumen de este atlas. Se realizó un análisis de correlación para encontrar las características que mostraron una relación con la clasificación de los pacientes en cada caso de estudio. Las características radiómicas, junto con información demográfica de los pacientes e información obtenida de la herramienta del software fueron utilizadas para entrenar diferentes algoritmos de aprendizaje autónomo y realizar predicciones en cada caso de estudio. También se realizaron pruebas utilizando solo las características correlacionadas con la clasificación de los pacientes en cada caso de estudio, para entrenar a los algoritmos y hacer nuevas predicciones.

La mayoría de de las características que mostraron una correlación con la clasificación de los pacientes en los tres casos de estudio, se midieron en regiones frontales, temporales y parietales. Sin embargo, algunas de las características con mayor correlación fueron medidas en el cerebelo, tálamo y cíngulo. De las predicciones utilizando todas las características para el entrenamiento, se obtuvieron valores de clasificación de entre 67 % y 87 % con los diferentes algoritmos, destacando el de red neuronal. En las pruebas utilizando las características con mayor correlación, se obtuvieron exactitudes de clasificación de entre 80 % y 93 %, destacando nuevamente el algoritmo de red neuronal. También se tuvieron valores de sensibilidad/especificidad de hasta 100 %/89 % utilizando las mejores características.

Se encontraron características radiómicas medidas en algunas regiones cerebrales, que muestran una distribución de valores que puede ser considerada diferente entre las dos clases de pacientes para los tres casos de estudio. Sin embargo, es necesario realizar más trabajos utilizando muestras con un mayor número de pacientes para poder establecer medidas de referencia que sirvan como apoyo en el diagnóstico de la EA. Los valores de rendimiento obtenidos utilizando características con mayor correlación, fueron superiores a los obtenidos utilizando todas las características de los pacientes, además de reducir los tiempos de entrenamiento de los algoritmos considerablemente, por lo que se deben utilizar estos métodos de selección de características en trabajos futuros. Los resultados de las predicciones son prometedores, pero es necesario realizar nuevos trabajos utilizando protocolos de adquisición de imágenes más controlados.

Abstract

Dementias are a group of diseases that mainly affect the elderly population and generally cause impairment of cognitive and intellectual abilities. One of the most frequently occurring dementias, mainly after the 60th, is Alzheimer's Disease (AD), which is a progressive neurodegenerative disease characterized by a loss of cognitive and memory functions. As part of the pathophysiology of the disease, one of the key metabolic changes observed in the Alzheimer's brain patients is a decrease in glucose metabolism. This change in glucose metabolism is a consequence of lesions in brain structures caused by the accumulation of β amyloid, a peptide that forms senile plaques, and tau protein. These lesions affect the regulation, absorption and utilization of glucose. Medical imaging techniques can be used as support in the diagnosis of AD to assess anatomical and physiological changes in patients' brains, with Positron Emission Tomography (PET) being the one which it is possible to study metabolic changes. Radiodiagnosis increasingly relies on computational image analysis tools that assist in medical decisions. One image analysis technique is radiomics, which involves obtaining data matrices from intensity and shape information in a region of interest. Radiomic features have demonstrated their utility in obtaining important information in many diseases; for example, in neurological diseases radiomics has been used to measure features in brain regions that have implications with a specific disease. Large amounts of data can be obtained by radiomic analyses, these data must be analysed by statistical methods and new data science techniques. Data science is one of the most important areas nowadays as many sectors, including health sciences, use large amounts of data that must be properly managed. This need has impulse the use of Machine Learning algorithms that allow identify data patterns, make predictions and even take a decision. In AD study, there are some works in the use of radiomic features measured in $[^{18}F]$ FDG-PET images to train machine learning (ML) algorithms and classify patients with and within the disease, perhaps more investigations using different methodologies to measure radiomic features and training or fitting a model are needed to use these parameters as a support in AD diagnosis.

A sample of 75 $[^{18}F]$ FDG-PET studies of patients with different diagnoses was used: 25 AD suspected, 25 with some neurological disease and; 25 healthy controls. All studies were retrospectively acquired from the PET/CT Unit at UNAM . Two studies were made considering the diagnosis of the patients as the target class: using patients with suspected AD and patients with some ND, and using patients with suspected AD and controls. In addition, a third study was made using patients with suspected AD and controls again, but considering the classification given by an Alzheimer's analysis tool of a commercial software as the target class. In each study there were 50 patients, all PET images were normalized to the HAMERS-N30R83 brain atlas space and then radiomic features were measured in each volume of this atlas. A correlation analysis was performed to find features that showed a relation with patients' classification in each study. Along with demographic information of the patients and information obtained from the softwa-

re tool, the radiomic features were used to train different ML algorithms and make predictions in each study. Another trial were also carried out using only features correlated with the classification of patients in each study, to train the algorithms and make new predictions.

Most of the features that showed a correlation with patients' classification in the three study cases were measured in frontal, temporal and parietal regions. However, some of the most correlated features were measured in cerebellum, thalamus and cingulate. From predictions using all of the features for training, accuracies between 67% and 87% were obtained, with the best performances by neural network. From the trials using only the most correlated features with patients' classification, accuracies between 80% and 93% were obtained, with the best performances again by neural network algorithm. Also, outcomes of sensibility/specificity of 100%/89% were obtained using the best features.

Radiomic features measured in some brain regions that show different distribution of values between the two classes of patients were found for the three studies. However, more studies using samples with a larger number of patients to establish reference values that serve as a support in AD diagnosis are needed. The performance results obtained using features with higher correlation were better than those obtained using all the patients' features, in addition less training times were needed, so these feature selection methods should be used in future works. The results of the predictions are promising, but it is necessary to carry out new works using more controlled image acquisition protocols.

Índice

1. Introducción	13
1.1. La enfermedad de Alzheimer	13
1.2. Tomografía por Emisión de Positrones	15
1.2.1. Bases de la Tomografía por Emisión de Positrones	15
1.2.2. Estudios PET con $[^{18}\text{F}]$ FDG	19
1.2.3. Medidas cuantitativas y semicuantitativas en estudios PET	20
1.3. Radiómica	22
1.3.1. El proceso de un estudio radiómico	23
1.3.2. Características radiómicas	24
1.4. Algoritmos de aprendizaje autónomo	25
1.4.1. Algoritmos de aprendizaje supervisado	26
1.4.2. Decision Tree	27
1.4.3. Random Forest	28
1.4.4. Regresión Logística	29
1.4.5. Máquinas de soporte vectorial	29
1.4.6. K vecino más cercano	31
1.4.7. Redes neuronales	32
1.4.8. Aplicación de modelos de aprendizaje autónomo en medicina	34
1.5. Justificación	36
1.6. Hipótesis	37
1.7. Objetivos	38
2. Metodología	39
2.1. Obtención de las imágenes PET	39
2.2. Obtención de mapas de t valores y clasificación	41
2.3. Preprocesamiento de imágenes PET	42
2.4. Segmentación de los volúmenes de interés	42
2.5. Adquisición de las características de radiómica	43
2.6. Análisis estadístico	44
2.7. Entrenamiento de los algoritmos	46
2.8. Pruebas y predicciones	46
2.9. Elección de características importantes y pruebas finales	48
3. Resultados	51
3.1. Información demográfica y características de pacientes	51
3.2. Imágenes $[^{18}\text{F}]$ FDG-PET	56

3.3. Mapas de t valores	56
3.4. Análisis de correlación	58
3.5. Prueba de Mann-Whitney U	62
3.6. Predicciones utilizando todas las características	63
3.7. Predicciones con las mejores características	65
3.8. Clasificación por clústers	67
4. Discusión	69
5. Conclusiones	76
6. Trabajo a futuro	78

Capítulo 1

Introducción

1.1. La enfermedad de Alzheimer

Uno de los problemas de salud pública más importantes actualmente son las demencias debido a que, en la mayoría de los casos, las personas que padecen algún tipo de demencia en etapas avanzadas son incapaces de valerse por sí mismas. Debido a esto, las personas que padecen algún tipo de demencia generalmente requieren cuidados especiales ya que estas enfermedades normalmente afectan a la población senil [1]. Las demencias son un grupo de enfermedades de carácter irreversible que dañan de manera estructural, toxicológica y fisiológica el cerebro y que se presentan como una alteración de las capacidades cognitivas e intelectuales que afecta las actividades de la vida diaria. Esta sintomatología se puede confundir con el deterioro cognitivo que se da por la edad avanzada, debido a esto, la Asociación Americana de Psiquiatría define el Diagnostic and Statistical Manual of Mental Disorders o DSM-IV con la finalidad de tener los criterios diagnósticos para la demencia [2]:

Criterio A. Desarrollo de déficits cognitivos múltiples que se manifiestan por:

1. Alteración de la memoria (alteración de la capacidad de aprender nueva información o recordar información previamente aprendida).
2. Una o más de las siguientes alteraciones cognitivas:
 - Afasia (deficiencia del lenguaje)
 - Apraxia (deficiencia motora)
 - Agnosia (deficiencia en el reconocimiento de objetos)
 - Alteración de la función ejecutiva (capacidad para el pensamiento abstracto y para planificar, iniciar, secuenciar, monitorizar y detener un comportamiento complejo).

Criterio B. Los defectos cognitivos de los criterios A1 y A2 han de ser lo suficientemente graves como para ocasionar un deterioro significativo de la actividad social o laboral y deben representar un déficit respecto a un nivel previo.

Clasificar las demencias no es tarea fácil, sin embargo, se pueden catalogar por el tipo de alteración proteica que las origina o por la sintomatología que ocasionan. Una de las demencias que

es posible identificar fácilmente en etapas avanzadas y que se presenta con mayor frecuencia es la Enfermedad de Alzheimer (EA), si en embargo, en las primeras etapas de la enfermedad se complica su identificación.

La Enfermedad de Alzheimer es un tipo de demencia que se presenta principalmente en la población de edad avanzada y que se caracteriza por un deterioro neurodegenerativo irreversible crónico que evoluciona hasta agravarse y ocasionar que quien la padece no pueda valerse por si mismo. La EA afecta cada vez a más personas a nivel mundial y en México existe una prevalencia del 7.3% y una incidencia de 27.3 (1000 personas/año) de la población adulta mayor [3], además, se espera que para el año 2050 la cifra de personas con EA en nuestro país alcance los 3.5 millones, lo que representará un problema importante para el sistema de salud.

La EA se distingue de las otras demencias principalmente porque se conservan las capacidades motoras, aún en etapas avanzadas, y porque el paciente presenta un deterioro de la memoria a corto plazo y también de la memoria a largo plazo en etapas avanzadas de la enfermedad. Las causas de la enfermedad pueden ser infecciones, problemas cerebrovasculares o hereditarios, pero la causa principal se asocia con una concentración anormal de las placas amiloide beta ($A\beta$) formadas por el polipéptido del mismo nombre, esto puede ocasionar diferentes tipos de atrofas en los lóbulos cerebrales ya que, debido a la incursión de dichas placas seniles, la sinapsis neuronal puede fallar y ocasionar la muerte de varias neuronas [4]. Debido a la neurodegeneración y disminución de la densidad sináptica ocasionadas por la acumulación del $A\beta$, sumado a otros factores que propician resistencia a la insulina, el metabolismo de glucosa se ve afectado considerablemente en pacientes con EA [5]. Además de la acumulación del $A\beta$, algunas características neuropatológicas de la enfermedad son la incursión intracelular de agregados τ hiperfosforilados (ovillos neurofibrilares o NFTs por sus siglas en inglés), atrofia cerebral, déficit colinérgico, neuroinflamación, disfunción sináptica y depleción celular [6]. Estas características se propagan en las regiones cerebrales durante años, afectando circuitos cognitivos. La enfermedad se considera como un espectro continuo ya que tiene como precursor un deterioro cognitivo leve o Mild Cognitive Impairment (MCI) que puede progresar hasta convertirse en Alzheimer [7]. Sin embargo, no todos los MCI progresan a la enfermedad ya que pueden existir diferentes tipos de deterioro cognitivo y los mecanismos por los cuales evoluciona a la EA aún se encuentran en estudio [8].

Además de las mutaciones en los genes involucrados con la expresión de la proteína precursora del $A\beta$, se han asociado otros factores de riesgo intrínsecos siendo la edad y el sexo los más determinantes, y extrínsecos como la dieta, el consumo de alimentos procesados, el tabaquismo y el consumo de alcohol, sin embargo, los estudios al respecto aun no son concluyentes ya que no se conoce muy bien la forma en la que estos factores propician la enfermedad [3]. Específicamente para la dieta, algunos estudios indican que las dietas altas en proteínas de origen animal (un adulto solo requiere del 10% al 15% de calorías por proteínas del total de su consumo) podrían ocasionar una disminución del sodio extracelular, lo cual ocasiona un estrés hiperosmótico en las células del sistema nervioso central y esto a su vez induce fosforilación y la producción de la proteína precursora del péptido $A\beta$ [2], la cual en concentraciones altas se considera como patológica ya que forma las placas $A\beta$ que se asocian con la EA.

El diagnóstico más certero de la enfermedad es el análisis histopatológico postmortem, sin embargo, para realizar un diagnóstico durante la enfermedad se pueden realizar diferentes pruebas, la más usada a nivel clínico es un test que evalúa el estado cognitivo del paciente. El problema es que esta técnica es subjetiva y depende de la experiencia del evaluador. Por lo anterior, se han desarrollado diversas técnicas auxiliares que ayudan a obtener un diagnóstico diferenciado como las técnicas de imagen funcional, siendo las más importantes la Resonancia Magnética Funcional (RMF) y la tomografía por emisión de positrones (PET, por sus siglas en inglés), mediante esta última se pueden realizar estudios con diferentes radiofármacos asociados a moléculas implicadas con la aparición de la EA, como lo son el péptido $A\beta$ o la glucosa [9]. Un radiofármaco es un fármaco que consiste en un radionúclido (núcleo radiactivo) asociado a una molécula biológica de interés [10]. En la figura 1.1 podemos observar un ejemplo de un estudio PET utilizado para asistir en el diagnóstico de la Enfermedad de Alzheimer.

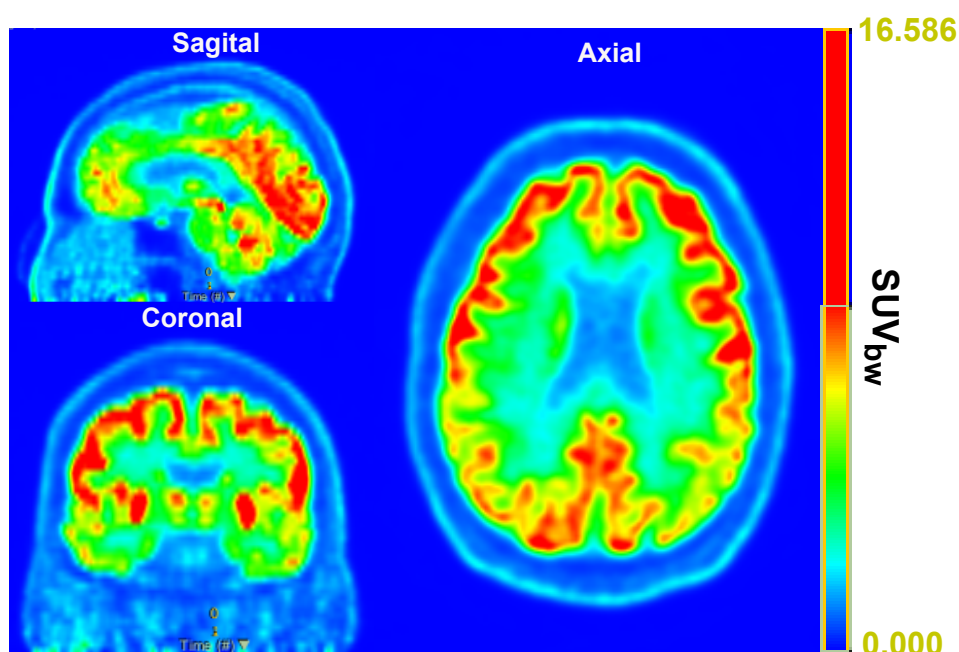


Figura 1.1: Ejemplo de un estudio PET con $[^{18}\text{F}]$ FDG para apoyo en diagnóstico de EA, tomado de la base de datos de la Unidad PET/CT de la Facultad de Medicina de la UNAM.

También existen otros métodos basados en la cuantificación de la proteína τ en el fluido craneoespinal, ya que puede ser un indicativo de la severidad de la enfermedad [11]. Sin embargo, el análisis del fluido para determinar la cantidad de τ aún está en discusión, además tiene la desventaja de ser invasivo ya que requiere de su extracción en la zona lumbar.

1.2. Tomografía por Emisión de Positrones

1.2.1. Bases de la Tomografía por Emisión de Positrones

La tomografía por emisión de positrones o PET por sus siglas en inglés, es una técnica de imagen de medicina nuclear ampliamente utilizada actualmente en países de alto ingreso y

cada vez más utilizada en México ya que permite obtener imágenes funcionales relacionadas con alguna molécula biológica de interés. Para esto se administra al paciente un radiofármaco, generalmente por vía intravenosa [10], con la finalidad de poder visualizar regiones de captación asociadas a un blanco molecular específico y relacionarlas con diferentes afecciones como lesiones o procesos tumorales. Se diferencia de la técnica de tomografía por emisión de fotón único o SPECT, por sus siglas en inglés, porque en dicha técnica se utiliza un fármaco asociado con un radioisótopo emisor de fotones gamma o de "fotón único", mientras que en el estudio PET se utilizan radioisótopos que emiten positrones, los cuales interactúan con los electrones de los átomos del medio, produciendo un fenómeno conocido como aniquilación. En este proceso, la masa del electrón y del positrón se transforma en energía en forma de dos fotones gamma de 0.511 MeV, los cuales son emitidos en direcciones opuestas a 180° [10] como se ilustra en la figura 1.2.

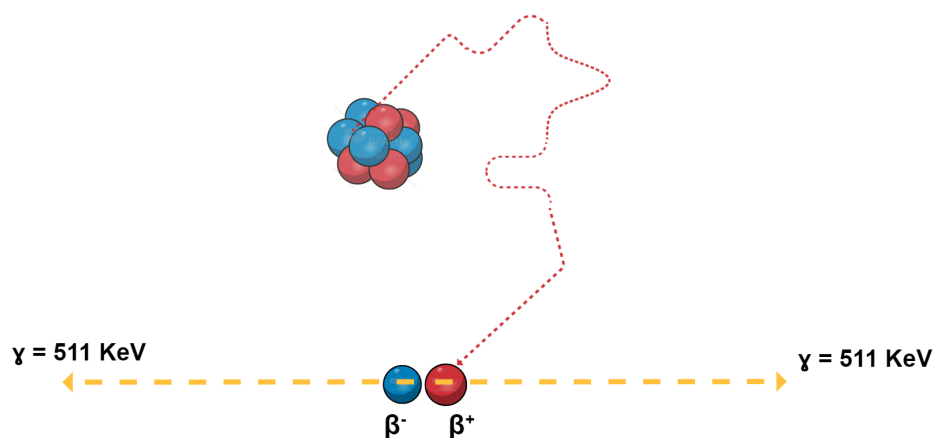


Figura 1.2: Proceso de aniquilación de un radionúclido emisor de β^+ .

Esto puede parecer un proceso complicado de utilizar como método de detección, sin embargo, esta característica permite obtener mayor precisión ya que, para que se considere un evento de aniquilación en un punto determinado como verdadero, se deben detectar en coincidencia los dos fotones gamma del proceso de aniquilación en dos detectores dentro de una ventana temporal y de energía, lo cual permite discriminar y eliminar fotones que no corresponden al proceso que se desea detectar y que normalmente son fotones dispersos o de otros procesos. Como su nombre lo indica, esta es una técnica tomográfica y por lo tanto requiere de un arreglo de detectores que cubra todas las direcciones de emisión del radiofármaco y permita realizar la reconstrucción de la imagen por corte, como se muestra en la figura 1.3.

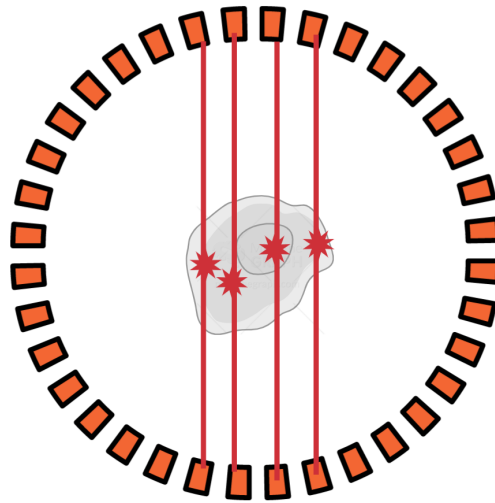


Figura 1.3: Ejemplo de un arreglo de detectores en un equipo PET.

Se observa que en la identificación de cada evento se forma una trayectoria recta virtual entre los dos detectores que detectaron el evento, esta recta se conoce como línea de respuesta o LOR, por sus siglas en inglés. Esta línea de respuesta, de manera ideal, permite ubicar la posición donde se produjo el evento de acuerdo con su tiempo de vuelo. El tiempo de vuelo o TOF por sus siglas en inglés, es el intervalo de tiempo que transcurre desde la producción del fotón, hasta su detección. Los dos fotones de aniquilación no serán detectados exactamente al mismo tiempo a menos que ocurran en el centro del campo de vista (FOV) y se tenga un medio uniforme, por lo tanto, esta diferencia temporal nos puede brindar información acerca de la posición del evento a lo largo de la LOR. Para asegurar que los fotones detectados provengan de un mismo evento, se establece una ventana temporal, la cual es un intervalo de tiempo corto (del orden de nanosegundos) que se establece cuando un fotón es detectado, y que se designa para que solo si otro fotón es detectado dentro de dicho intervalo, el evento será aceptado y se generará una línea de respuesta entre los detectores correspondientes, este proceso se conoce como colimación electrónica y permite evitar la detección de fotones dispersos del medio. En un equipo PET también se define una ventana energética que típicamente va de los 350 a los 650 KeV y de igual forma sirve para eliminar fotones que no corresponden a un evento de aniquilación, sino a fotones dispersos que han perdido de manera considerable su energía en el medio. A pesar de estas restricciones, aun hay un porcentaje importante de fotones dispersos que aumentan el ruido en una imagen PET y que por lo tanto afectan la resolución espacial. Además, existen otros factores que contribuyen a empeorar la resolución espacial del equipo como los métodos de reconstrucción, el tamaño del detector, la codificación, la incertidumbre en la co-linealidad de los dos eventos, la profundidad de interacción del fotón gamma en el detector y el alcance de los positrones, los cuales recorren cierta trayectoria luego del decaimiento β^+ del radionúclido y posteriormente ocurre el evento de aniquilación [10], como se ejemplificó en la figura 1.2.

Otro factor importante a considerar en estudios PET para diagnóstico es el radionúclido utilizado en cada radiofármaco, ya que este determinará la calidad de imagen de acuerdo con el alcance de los positrones que emite y la molécula biológica determinará la ruta metabólica que seguirá

el radiofármaco. El radiofármaco más utilizado en estudios PET es el [^{18}F] Fluorodesoxiglucosa ([^{18}F]FDG), el cual consiste en una molécula análoga de glucosa enlazada químicamente con un radionúclido de ^{18}F , el cual es un emisor de positrones [10] con vida media de aproximadamente 109 minutos. El ^{18}F se produce en aceleradores de partículas tipo ciclotrón, los cuales permiten acelerar un haz de protones a velocidades aproximadas a un tercio de la velocidad de la luz, para después hacerlo impactar con un blanco de ^{18}O generando la reacción protón-neutrón $^{18}\text{O}(p,n)^{18}\text{F}$ [12]. Los átomos de ^{18}F obtenidos se asocian químicamente con un precursor y luego de pasar por diferentes procesos se obtiene el análogo de glucosa [^{18}F]FDG [12]. La razón por la que se utiliza un radiofármaco que contiene un análogo de glucosa es por su importancia biológica en múltiples funciones como los procesos de desarrollo tumoral, ya que las células cancerosas requieren mayores cantidades de glucosa para la replicación celular y también aumentan considerablemente su consumo de glucosa cuando sufren daños luego de tratamientos como radioterapia o quimioterapia. La glucosa también es relevante en otros procesos fisiológicos, por ejemplo, en el cerebro se requiere una cantidad importante de glucosa en la mayoría de las regiones y esta cantidad va disminuyendo de manera natural conforme aumenta la edad, pero también puede verse afectada su captación por enfermedades que ocasionen daños fisiológicos y estructurales. Sin embargo, como ya se mencionó, existen otros biomarcadores que pueden ser útiles dependiendo del tipo de lesión o enfermedad que se presente y que pueden ser útiles para evaluar la captación de otro tipo de moléculas biológicas asociadas a algún proceso relacionado con la lesión o enfermedad. Por ejemplo, para localizar lesiones o tumores cerebrales, debido a que en esta región se tiene una alta captación de glucosa en la materia gris y moderadamente alta en la materia blanca [13], provocando que se tenga un fondo de captación elevado, lo cual en algunos casos puede complicar la identificación de lesiones cerebrales. Sin embargo, desde la llegada del PET se han venido implementando nuevos radiofármacos como la c-metionina, la cuál es muy útil para detectar tumores en el cerebro, ya que esta es captada en los procesos de división celular debido a que la metionina participa en este proceso, permitiendo de esta manera poder identificar una malformación cerebral por ejemplo, sin que exista captación en las otras zonas [14]. Para el estudio de la EA también existen estudios PET con radiofármacos específicos asociados con la deposición del $A\beta$, siendo el [^{11}C]Pittsburgh el más utilizado por tener buena sensibilidad y especificidad [15]. Sin embargo, debido a la vida media corta del ^{11}C (20 minutos), se han desarrollado radiofármacos utilizando ^{18}F como radionúclido y cuyo blanco molecular sea el $A\beta$, siendo el [^{18}F]flutemetamol, el [^{18}F]florbetapir y el [^{18}F]florbetaben los únicos aprobados por la FDA para su uso clínico [6]. Los radiofármacos asociados con la deposición del $A\beta$ normalmente son utilizados en etapas tempranas de la enfermedad cuando se tiene sospecha o cuando se tienen síntomas de MCI [16]. Además, en el ámbito de la investigación se encuentran en estudio radiofármacos asociados a la presencia de la proteína τ , a procesos inflamatorios y a moléculas relacionadas con la vía colinérgica [11].

El PET por sí mismo, tiene numerosas ventajas a comparación de otros estudios de imagen como la tomografía computarizada (CT) o la resonancia magnética (RM) ya que, a pesar de no obtener información anatómica, se obtiene una imagen funcional que permite evaluar procesos biológicos importantes como el correcto funcionamiento de un órgano en específico, los procesos de reparación de algún daño en tejido o el metabolismo de las malformaciones e información acerca de su densidad [17]. Otra ventaja importante de los estudios PET sobre las otras técnicas de imagen es que es posible llevar un seguimiento del progreso de la enfermedad en estudio e identificar de inmediato si se tiene una respuesta positiva al tratamiento de dicha enfermedad,

permitiendo tomar decisiones anticipadas acerca de la forma en la que se procederá a lo largo del tratamiento.

1.2.2. Estudios PET con $[^{18}\text{F}]$ FDG

Como se mencionó, el radiofármaco más utilizado en estudios PET es el $[^{18}\text{F}]$ FDG debido a su alta disponibilidad y versatilidad en este tipo de estudios. Para producir este radiofármaco se requiere de un precursor químico llamado triflato de manosa, el cual es un análogo de glucosa, que es ligado a un átomo de Flúor-18 y pasa por diversos procesos fisicoquímicos para obtener el $[^{18}\text{F}]$ FDG [18]. La disponibilidad del precursor químico, la vida media del radionúclido y la experiencia en la síntesis de FDG son los factores más importantes que explican que sea el radiofármaco más utilizado en estudios PET. Al ser un análogo de glucosa, este radiofármaco penetra la barrera celular principalmente mediante transportadores de glucosa 1 o GLUT-I, los cuales permiten llevar a la glucosa del plasma sanguíneo al citoplasma celular. Después de este proceso, la glucosa normalmente es fosforilada por una enzima llamada hexoquinasa y después puede seguir dos rutas metabólicas, ya sea para producir ATP o para almacenarse como glucógeno. Sin embargo, cuando el $[^{18}\text{F}]$ FDG es fosforilado por la enzima, este no sigue ninguna de las rutas metabólicas debido a que contiene un átomo de ^{18}F en la posición del carbono 2, por lo que solo permanece fosforilado [19] como se observa en la figura 1.4.

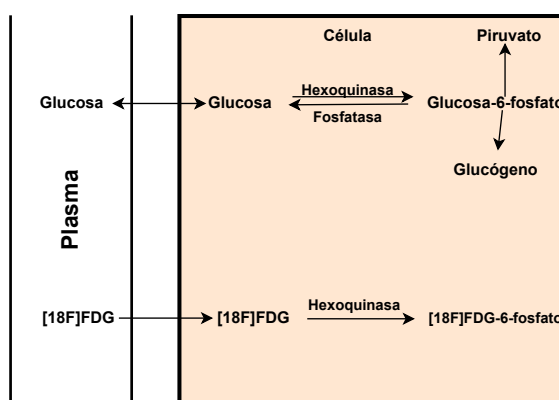


Figura 1.4: Diagrama de la ruta metabólica del $[^{18}\text{F}]$ FDG y glucosa.

No obstante, la captación del $[^{18}\text{F}]$ FDG se puede considerar como una medida del metabolismo de glucosa debido a que su presencia es directamente proporcional a la expresión de la enzima hexoquinasa y por lo tanto, a la tasa de metabolización de glucosa del tejido [20]. El metabolismo de glucosa es relevante en múltiples procesos del cuerpo siendo la fuente principal de energía, por ejemplo, en la función muscular. También es relevante en procesos oncológicos debido al alto consumo energético que requiere la replicación celular. El metabolismo de glucosa también es muy importante en el cerebro, ya que el 95 % del ATP requerido para las funciones cerebrales proviene de la glucosa, por lo que gran parte de esta es transportada al cerebro [21]. Debido a esto, el metabolismo de la glucosa está ampliamente relacionado con la función neuronal y sináptica [6], ya que las neuronas no se replican y por lo tanto no requieren energía para este fin. Esto vuelve a los estudios PET con $[^{18}\text{F}]$ FDG un biomarcador importante de actividad neuronal y densidad sináptica. Este tipo de estudios cobra relevancia en el análisis de demencias en las que se ve afectado el metabolismo de glucosa, por ejemplo, la Enfermedad

de Alzheimer se relaciona con un patrón hipometabólico de glucosa en el cerebro [6]. Sin embargo, no es sencillo identificar la presencia de la EA ya que el consumo de glucosa cerebral disminuye de manera natural por la edad, debido a esto, se han estudiado los patrones de hipometabolismo en las diferentes zonas cerebrales que son característicos de dicha enfermedad, encontrando que en etapas tempranas de la EA se reduce el metabolismo de glucosa en el giro del cíngulo, el corte temporoparietal y la precuña, y en etapas más avanzadas se reduce en el lóbulo occipital y frontal [20]. Esto vuelve a los estudios $[^{18}\text{F}]\text{FDG-PET}$ un auxiliar necesario en el diagnóstico de la enfermedad y también cobra relevancia la cuantificación de la captación de FDG para establecer umbrales.

1.2.3. Medidas cuantitativas y semicuantitativas en estudios PET

A pesar de las ventajas que brindan las imágenes obtenidas en los estudios PET, actualmente el diagnóstico está basado principalmente en la inspección visual y cualitativa, lo cual implica que el estudio depende mucho de la experiencia del médico radiólogo y que la información cuantitativa que contiene la imagen no se aproveche como herramienta para asistir el diagnóstico. Esto hace que la cuantificación de radiofármacos en estudios de imagen PET sea una de las vertientes de investigación más importantes, ya que la finalidad es determinar el metabolismo de la molécula biológica que se analiza. Para esto se han desarrollado modelos matemáticos que utilizan tejidos de referencia para comparar su concentración de actividad con la de los tejidos de interés, estos se conocen como modelos compartimentales [22]. Los modelos compartimentales son modelos matemáticos basados en la representación de los diferentes procesos farmacocinéticos como compartimentos, tal y como se ilustra en la figura 1.5.

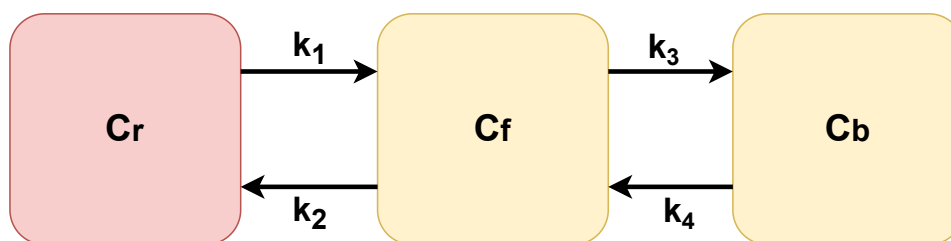


Figura 1.5: Representación de los compartimentos supuestos en los modelos compartimentales.

Estos compartimentos pueden representar tejidos o procesos por los que pasa el fármaco en estudio, normalmente el C_r representa el compartimento vascular, es decir, cuando el fármaco se encuentra en el torrente sanguíneo y los compartimentos C_f y C_b representan diferentes estados o procesos metabólicos por los que pasa el fármaco dentro del tejido de interés. Estos modelos tienen la finalidad de estimar microparámetros que representan la cuantificación de la molécula que se desea estudiar, a partir de la determinación de macroparámetros. La suposición de compartimentos permite modelar el proceso de captación de un radiofármaco en un tejido de

interés a partir de ecuaciones diferenciales dadas por su farmacocinética

$$\begin{aligned}\frac{dC_r(t)}{dt} &= k_2C_f(t) - k_1C_r(t) \\ \frac{dC_f(t)}{dt} &= k_1C_r(t) - (k_2 + k_3)C_f(t) + k_4C_b(t) \\ \frac{dC_b(t)}{dt} &= k_3C_f(t) - k_4C_b(t)\end{aligned}\quad (1.1)$$

Estas ecuaciones permiten relacionar la captación en los compartimentos C_i , a un tiempo t , teniendo cada compartimento su tasa de cambio de entrada o salida k_i . La finalidad de estos modelos es hallar las tasas de cambio para conocer la captación real del radiofármaco en el tejido de interés, sin embargo, se tienen muchas incógnitas en el sistema de ecuaciones y esto complica su resolución, por lo que en su lugar se obtienen soluciones globales que contienen a estas tasas de cambio denominadas macroparámetros. Para obtenerlos existen diferentes formas de simplificar el modelo, dependiendo de las consideraciones que se hagan y del tipo de proceso de captación (reversible o irreversible), la mayoría se basa en la suposición de un tejido de referencia y en métodos de análisis gráfico. En general, los procesos de captación reversibles se refieren a aquellos en los que la molécula biológica de interés puede ingresar y salir nuevamente del tejido de interés, mientras que en los procesos irreversibles la molécula se mantiene dentro de este tejido. Uno de los modelos que considera procesos reversibles es el modelo de tejido de referencia simplificado o SRTM, por sus siglas en inglés, y se basa en la misma consideración de un tejido de referencia pero aplicando el siguiente cambio de variable [23]

$$k_1/k_{2a} = (k_1/k_2) \cdot (1 + BP) \quad (1.2)$$

En donde BP es el macroparámetro denominado potencial de unión, resolviendo el sistema de ecuaciones diferenciales se obtiene la siguiente solución de la concentración de actividad en el tejido de interés

$$\begin{aligned}C_t(t) &= R_1C_r(t) \\ &+ \{k_2 - R_1k_2/(1 + BP)\} C_r(t) * \exp\{-k_2t/(1 + BP)\}\end{aligned}\quad (1.3)$$

En donde las demás constantes son parámetros de ajuste. Para procesos irreversibles, es posible emplear otro tipo de modelos que toman en cuenta esta característica, por ejemplo el modelo de Patlak [24], el cual de igual forma considera dos compartimentos y se basa en el cálculo del cociente de la concentración de actividad en la región de interés, entre la concentración de actividad en la región de referencia, matemáticamente se puede expresar de la siguiente manera

$$\frac{ROI(t)}{P(t)} = K_i \frac{\int_0^t P(t) dt}{P(t)} + V_B \quad (1.4)$$

En donde $ROI(t)$ es la concentración de actividad en la región de interés y $P(t)$ en la región de referencia. Esta ecuación tiene la forma de la ecuación de una recta con pendiente K_i , por lo que al realizar un ajuste lineal se puede obtener este macroparámetro. Al realizar un estudio PET, normalmente se toman medidas semicuantitativas en la práctica clínica, que en teoría son representativas de los macroparámetros o medidas cuantitativas. No obstante las medidas cuantitativas se siguen midiendo a través de estudios dinámicos que permiten obtener curvas de actividad contra tiempo y realizar ajustes con los modelos compartimentales. Esto generalmente

se realiza con fines de investigación, pero también para determinar el tiempo óptimo que se debe esperar después de inyectar un radiofármaco para iniciar con la adquisición y tomar las medidas semicuantitativas [25]. El parámetro semicuantitativo por excelencia de los estudios [^{18}F]FDG-PET es el valor de captación estandarizado o SUV, el cual relaciona la captación del radiofármaco en una región en específico con la actividad total suministrada y con el peso del paciente y se calcula de la siguiente manera

$$SUV = \frac{\hat{A}}{(A_T/w)} \quad (1.5)$$

En donde \hat{A} representa la concentración de actividad en la región de interés (ROI, por su siglas en inglés), A_T representa la actividad total inyectada al paciente y w representa el peso del paciente [26]. Como podemos observar, el numerador en la fórmula se relaciona con la captación en la región de interés o píxeles de interés en la imagen y el denominador se relaciona con la actividad específica inyectada al paciente, por lo que indirectamente se tiene un parámetro que permite cuantificar la captación del radiofármaco en cada píxel de la imagen. Del SUV se pueden obtener otras medidas que permiten reducir errores debidos a la resolución espacial limitada del estudio y a otros errores relacionados con el procesamiento y la electrónica. Los más importantes son el SUV máximo (SUV_{max}) y el SUV promedio (SUV_{mean}). Para obtenerlos primero se debe definir una región de interés o ROI la cual debe delimitar la zona de la lesión o tejido en estudio. De esta manera, el SUV_{max} representa el valor SUV máximo que se tiene en la ROI, este valor corresponde a alguno de los voxels de la región. El SUV_{max} es uno de los parámetros más usados en la práctica clínica por su utilidad en lesiones pequeñas, debido su poca dependencia con el efecto parcial de volumen. Sin embargo, este valor puede verse afectado por el ruido de la imagen debido a su proceso reconstrucción. El SUV_{mean} en cambio, es un valor más representativo de toda la región en estudio, pero puede variar dependiendo de qué voxels son incluidos en la ROI [27]. Un parámetro clasificado como un subtipo de SUV_{mean} es el SUV_{peak} o SUV pico, en el cual se tiene una ROI específicamente bien definida y en esta se incluyen los valores de vóxel más altos (incluyendo al voxel con el SUV_{max}) para medir el promedio de estos valores, asegurando la reproducibilidad de las mediciones y reduciendo la incertidumbre por ruido [28]. Estas medidas semicuantitativas brindan mayor información acerca del metabolismo de la molécula blanco en un estudio PET y permiten establecer valores de referencia. Sin embargo, existen otras características físicas de las lesiones en estudio, en las cuales las medidas semicuantitativas convencionales no pueden brindarnos la información necesaria.

1.3. Radiómica

La radiómica es una técnica de análisis de imagen que se encarga de entrar más a fondo en la búsqueda de parámetros que no se identifican a simple vista y que pueden ser de importancia para caracterizar correctamente el problema que se tiene. En general, la radiómica consiste en la conversión de imágenes en diferentes matrices de datos extraíbles que pueden ser analizados [29], estos datos se obtienen con la finalidad de definir características que estén relacionadas con el metabolismo, fisiología, anatomía o algún otro proceso biológico de la enfermedad o problema en cuestión. Cabe aclarar que las imágenes por sí mismas son matrices de datos y lo que se obtiene de estas son otras matrices que servirán como descriptores de diversas características

de la imagen. La radiómica puede ser una alternativa para algunos procedimientos invasivos de diagnóstico, brindando información acerca de la fisiología del tejido mediante el estudio de imagen médica adecuado.

Los campos de estudio explorados por la radiómica que han tenido una mayor evolución en los últimos años son el radiológico y el oncológico, esto se ve reflejado en un mayor número de investigaciones en esas áreas [30]. La relevancia de la radiómica se debe a que se pueden definir parámetros que permitan identificar características importantes de diferentes tipos de lesiones tales como tamaño, su forma, densidad, heterogeneidad, textura, características superficiales e internas y su metabolismo. Todas y cada una de estas características son de suma importancia no solo para identificar el tipo de tumor que se tiene, sino para determinar su grado de malignidad o agresividad, haciendo posible no solo un diagnóstico, sino también un pronóstico del desarrollo de la enfermedad. No obstante, el cáncer no es el único grupo de enfermedades en los que la radiómica es relevante, ya que existen otras enfermedades como las demencias en donde la radiómica puede ser una herramienta importante, debido a que en muchas ocasiones es complicado dar un diagnóstico de este tipo de enfermedades cuando aún se encuentran en etapas tempranas o incluso durante el desarrollo de las mismas y los procedimientos invasivos en esta zona son aún más complicados y arriesgados, por lo que en muchos casos la confirmación del diagnóstico del paciente llega hasta que se hace una inspección postmortem de la zona, esto ocurre por ejemplo en la EA.

La relevancia de la implementación de la radiómica es clara, sin embargo, aún es un área de estudio relativamente reciente y en desarrollo que puede o no ser considerada para su regulación e implementación en la práctica clínica.

1.3.1. El proceso de un estudio radiómico

El proceso que se sigue en radiómica consta principalmente de cuatro etapas: Adquisición de las imágenes médicas, segmentación del volumen de interés (VOI, por sus siglas en inglés), extracción de características y análisis de los datos obtenidos. El estudio de imagen médica puede ser de cualquier modalidad, dependerá de la naturaleza de la lesión o tumor y de qué estudio puede brindar más y mejor información. Para estudios funcionales o metabólicos, normalmente se emplean estudios PET simples o en conjunto con CT cuando se requiere visualizar también información anatómica, también existen combinaciones de PET con RM.

Antes de realizar la extracción de características radiómicas, se debe definir un método de segmentación adecuado para delimitar el volumen de interés, ya sea de manera automática o manual [31], también lo puede realizar el usuario con ayuda de programas para la segmentación semiautomática. La segmentación consiste principalmente en operaciones para dividir una imagen en las partes que la componen, también se puede definir como la separación de objetos que se encuentran en la imagen [32]. Para esto, se pueden emplear diferentes métodos matemáticos que se pueden resumir en dos categorías:

- Métodos de umbralización
- Métodos de identificación de bordes

De manera general, los métodos de identificación de bordes se basan en la discretización de funciones que tienen un perfil más continuo, para identificar los bordes correspondientes al objeto de interés [32]. En cambio los de umbralización, son más útiles para identificar regiones en las que el volumen de interés no se encuentra muy bien definido y se establece un valor de intensidad de píxel como umbral para delimitar la región. La finalidad de la segmentación es hallar los volúmenes de interés, los cuales dependiendo de la enfermedad en estudio, representan las regiones relevantes en los procesos fisiológicos de la enfermedad o los tejidos en donde se encuentran las lesiones en estudio y por lo tanto son determinantes para su diagnóstico. De estos volúmenes de interés se obtienen las características radiómicas mediante métodos matemáticos para calcular matrices de correlación de niveles de gris. Luego de realizar un análisis adecuado, las características radiómicas medidas en el VOI pueden ser correlacionadas con la enfermedad o afección en estudio.

1.3.2. Características radiómicas

Las características radiómicas se pueden clasificar de diferentes maneras, por ejemplo, de acuerdo a su dimensión, su naturaleza (características de intensidad, de forma, etc.) o su finalidad. Por ejemplo, se pueden clasificar como de primer o segundo orden dependiendo si se trata de índices o de matrices [33]:

Características de primer orden

- Índices de forma
- Características de histograma de primer orden
- Índices convencionales
- Índices discretizados

Características de segundo orden

- Matriz de co-ocurrencia de nivel de gris (GLCM, por sus siglas en inglés)
- Matriz de dependencia del nivel de gris vecino (NGLDM, por sus siglas en inglés)
- Matriz de longitud de carrera de nivel de gris (GLRLM, por sus siglas en inglés)
- Matriz de longitud de la región de nivel gris (GLZLM, por sus siglas en inglés)
- Matriz de Zona de Tamaño de Nivel de Gris (GLSZM, por sus siglas en inglés)
- Matriz de Zona de Distancia de Nivel de Gris (GLDZM, por sus siglas en inglés)
- Matriz de Diferencia de Tono de Gris del Vecindario (NGTDM, por sus siglas en inglés)

Cabe aclarar que en este listado se mencionan familias de características y por cada familia, ya sea de matrices o de índices, se pueden obtener diferentes medidas. La definición y fundamento matemático de las características radiómicas utilizadas se puede consultar en los anexos de este trabajo.

Para la etapa del análisis de los datos obtenidos, existe una amplia gama de métodos de evaluación y no siempre se siguen las mismas metodologías, pero es recomendable aplicar lo que se ha comprobado que tiene resultados fiables. No obstante, debido a que se manejan grandes compendios de datos en las mediciones radiómicas, es necesario evaluar la posibilidad de aplicar métodos de ciencia de datos y análisis estadístico. Afortunadamente, la ciencia de datos es una de las áreas con mayor crecimiento actualmente, por lo que se tiene una amplia gama de métodos computacionales con los que se pueden obtener resultados prometedores, como se verá en secciones siguientes.

1.4. Algoritmos de aprendizaje autónomo

Los algoritmos de aprendizaje autónomo o Machine Learning son una serie de algoritmos que permiten a un programa aprender algo sin que esto esté explícitamente programado mediante reconocimiento de patrones. Para esto se requiere entrenar el modelo con datos de prueba que le permitan aprender la tendencia que siguen y predecir un resultado a partir de dicho entrenamiento [34]. En el proceso que se sigue al estudiar este tipo de problemas se debe considerar que los parámetros de cálculo se utilicen adecuadamente y dependiendo del caso que se analiza, es decir, que en cada modelo se deben elegir los parámetros con los que se obtienen los mejores resultados. Además, también es importante la elección de los datos que son relevantes para la toma de decisiones del modelo. En la figura 1.6 se resume el proceso que se sigue en un estudio utilizando un algoritmo de aprendizaje autónomo.

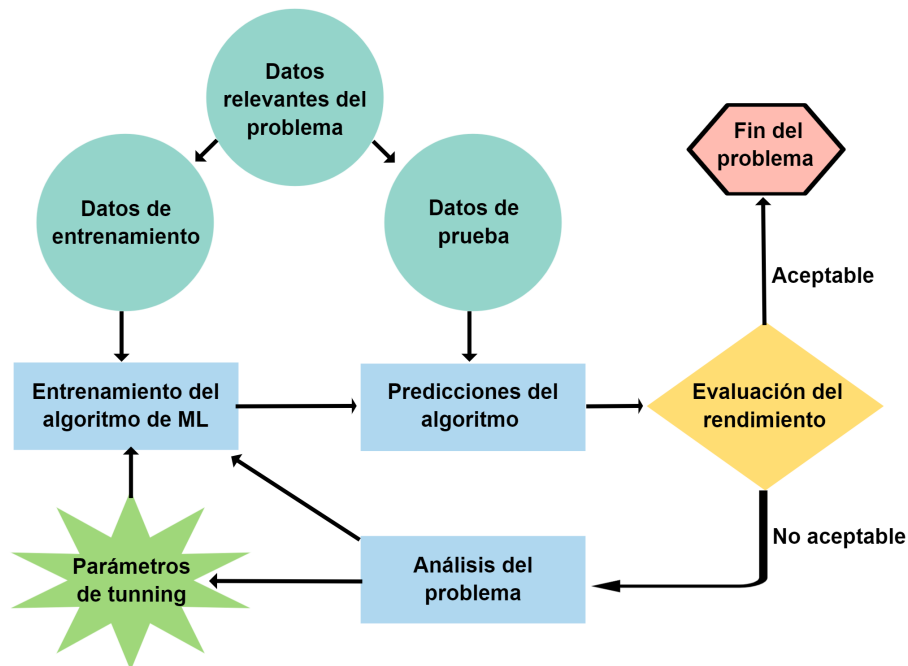


Figura 1.6: Esquema general del proceso que sigue un algoritmo de aprendizaje autónomo.

A pesar que existen varios tipos de modelos, a grandes rasgos los modelos de aprendizaje autónomo generalmente se dividen en supervisados y no supervisados. Los modelos de aprendizaje supervisado son los que requieren de cierta intervención externa, concretamente se requiere tener previamente la clasificación de cada instancia, para obtener una predicción del modelo. Para esto se deben emplear datos de entrenamiento que generalmente representan la mayor parte de la muestra y datos de prueba que generalmente representan una muestra más pequeña que los datos. Mediante el entrenamiento previo recibido, el modelo realizará una predicción a partir del patrón que identifique en los datos de prueba. El aprendizaje no supervisado por el contrario, no requiere de esta intervención externa, es decir, que no se requiere conocer la clasificación de las instancias previamente. En este caso los algoritmos aprenden los patrones y características importantes de los datos por sí mismos y utilizan estas características para clasificar cada instancia [35]. Se utilizan principalmente para la agrupación de datos en clústeres y la reducción de características.

Este tipo de algoritmos son de mucha utilidad para el manejo de grandes cantidades de datos y de hecho el aumento en la cantidad de datos permite que el modelo obtenga mejores resultados al realizar las predicciones, por lo que son una herramienta importante actualmente ya que la tendencia de los problemas es mejorar la precisión y exactitud de los resultados, lo cual en ocasiones conlleva aumentar la cantidad de datos para su análisis por la complejidad del problema.

1.4.1. Algoritmos de aprendizaje supervisado

Los algoritmos de machine learning, de manera general, siguen un mismo camino al analizar los patrones de un grupo de datos para aprender su comportamiento y realizar estimaciones. Específicamente los algoritmos de aprendizaje supervisado se distinguen porque los datos analizados contienen la solución deseada o esperada, estas soluciones se conocen como clases objetivo o targets y permiten al algoritmo no solo aprender el patrón de comportamiento de los datos de cada clase, sino también evaluar las predicciones (clasificación de las instancias) realizadas por el algoritmo, verificando si estas coinciden con la clase real de cada instancia. Debido a esto, los algoritmos de aprendizaje supervisado son utilizados principalmente para tareas de clasificación, ejemplo de ello son los algoritmos empleados para clasificar los diferentes tipos de correos electrónico recibidos, ya que el algoritmo se encuentra previamente entrenado para clasificar los correos no deseados e incluso sigue aprendiendo de acuerdo a las preferencias del usuario cuando le indicamos si un correo es o no deseado. También se pueden emplear para predecir o estimar un valor numérico a partir de un grupo de datos con características específicas denominadas predictores [34]. La principal diferencia con los algoritmos de aprendizaje no supervisado es que en estos últimos el grupo de datos a analizar no contiene las clases objetivo y, por lo tanto, el algoritmo únicamente analiza y predice el patrón que siguen los datos. Los algoritmos de aprendizaje supervisado más importantes son los siguientes:

- Decision tree
- Random forest
- K vecino más cercano (KNN, por sus siglas en inglés)
- Redes neuronales

- Regresión lineal
- Regresión logística
- Máquinas de soporte vectorial (SVM, por sus siglas en inglés)

A pesar de ser algoritmos de aprendizaje supervisado y de tener una metodología similar, cada uno realiza un análisis diferente del grupo de datos, con diferentes operaciones matemáticas y diferente lógica, algunos se basan en probabilidades y otros utilizan algoritmos estocásticos. La elección de alguno de estos algoritmos dependerá de las necesidades del problema, por ejemplo si se requiere realizar tareas de clasificación o regresión, si se tiene una muestra de datos pequeña, etc. En las siguientes secciones se explicará el fundamento general de algunos de los algoritmos y su aplicación.

1.4.2. Decision Tree

Uno de los algoritmos de machine learning más importante y más utilizado es el Decision tree o árbol de decisión, esto debido a su simplicidad y facilidad de uso y a los buenos resultados que se pueden obtener si se utiliza de manera correcta. Este algoritmo se puede emplear para realizar tareas de clasificación, es decir, para predecir la clase o subgrupo al que pertenece un dato de un grupo de datos de acuerdo a sus características. También es capaz de realizar tareas de regresión, es decir, predecir el valor numérico de un dato dependiendo de nueva cuenta de sus características. De manera general, un algoritmo tipo Decision Tree tiene una lógica similar a la que se muestra en la figura 1.7.

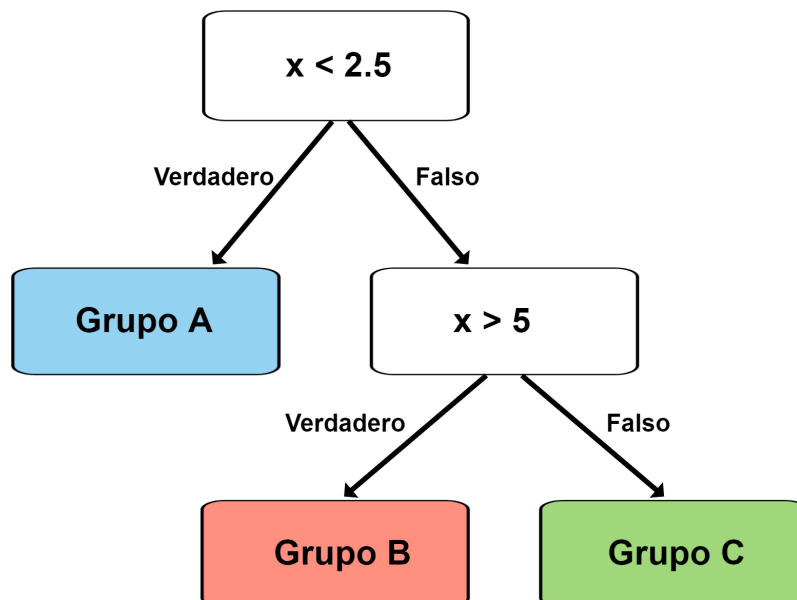


Figura 1.7: Ejemplo de la lógica de un algoritmo tipo Decision Tree.

A simple vista puede parecer que el algoritmo clasifica los datos del grupo dependiendo de su característica, sin embargo, lo importante a destacar es que el algoritmo por sí mismo toma las características más importantes del grupo de datos y las asocia a un grupo o clase determinada, dividiendo el total de datos originales en subgrupos. Después de ser entrenado, el algoritmo puede clasificar un nuevo grupo de datos de prueba y predecir a qué grupo pertenecen [34]. Se puede observar que en cada cuadro se tienen diferentes características con sus respectivos atributos (valores), estas características pueden ser definidas para cada modelo en específico por el usuario y pueden ser determinantes en el resultado de las predicciones. Otro factor a considerar es la profundidad del modelo, es decir, que tantas decisiones debería tomar para poder clasificar los datos, esto también puede ser definido por el usuario [34]. A pesar de ser útil cuando se tienen muestras pequeñas, este algoritmo toma ventaja cuando se cuenta con grandes cantidades de instancias y características.

1.4.3. Random Forest

El problema con los algoritmos tipo Decision tree es su naturaleza estocástica, ya que al realizar un nuevo proceso de entrenamiento y pruebas aún sin modificar los datos, los resultados obtenidos no siempre serán los mismos. Esto puede ser un problema sobretodo para los estudios en los que no se tiene un gran número de datos y las fluctuaciones probabilísticas pueden ser altas. Para resolver este problema, se han propuesto algoritmos que permitan tomar como base otros algoritmos más sencillos como el Decision tree e integrar todos los resultados obtenidos de diferente manera, por ejemplo tomando el resultado más frecuente de múltiples predicciones realizadas o tomando un valor promedio. Ese tipo de algoritmos se conocen como ensembles y se pueden entender como algoritmos compuestos de las clasificaciones realizadas por un grupo de algoritmos de aprendizaje supervisado. Los ensembles toman las clasificaciones dadas por el grupo de algoritmos para una instancia y la clasifican según la clase que tiene una mayor ocurrencia, como se ejemplifica en la figura 1.8.

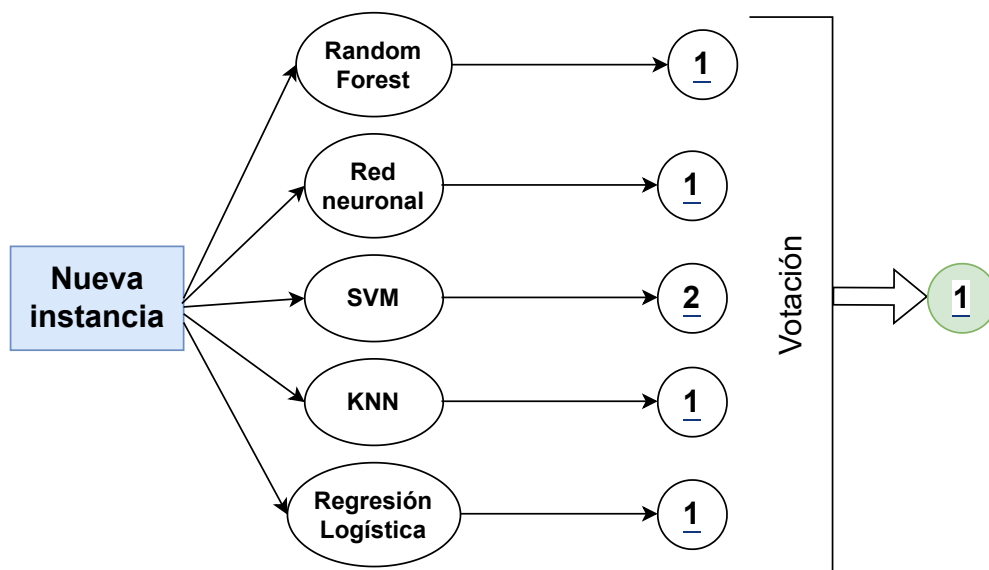


Figura 1.8: Ejemplo de la lógica de un algoritmo tipo ensemble.

Específicamente para los algoritmos tipo Decision tree, se pueden crear múltiples arboles que tomen decisiones luego de ser entrenados y realizar las pruebas con los mismos grupos de datos, este tipo de algoritmos compuestos se conocen como Random Forest [34]. Los algoritmos tipo Random Forest son ampliamente utilizados debido a que, al igual que los Decision tree, son fáciles de emplear y de entender, las predicciones son rápidas a pesar de la complejidad que conlleva tener varios árboles y los resultados suelen ser favorables. En general se tienen las mismas limitaciones que en los árboles de decisión en cuanto a la profundidad de los árboles, pero se reducen las imprecisiones debidas a la naturaleza estocástica de los mismos. Al igual que en los árboles de decisión, los Random Forest son útiles cuando se cuenta con grandes cantidades de instancias y/o características.

1.4.4. Regresión Logística

La regresión logística es un modelo de predicción binario que permite clasificar un grupo de datos en dos clases, generalmente denominadas objetivo, a partir de estimar la probabilidad de pertenencia a alguna de las clases. Esta probabilidad se puede estimar mediante el entrenamiento o ajuste de las características pertenecientes al grupo de datos en estudio, la expresión matemática vectorizada que describe la estimación de probabilidad se muestra en la ecuación 1.6 [34]

$$\hat{p} = h_{\theta}(\mathbf{x}) = \sigma(\mathbf{x}^T \boldsymbol{\theta}) \quad (1.6)$$

En donde \mathbf{x} representa las instancias (sujetos, objetos de estudio, etc.), $\boldsymbol{\theta}$ es un vector paramétrico y σ es la función logística normalizada. Esta última es el parámetro importante en el modelo, ya que es la función que permite obtener las probabilidades de las instancias de pertenecer a alguna de las clases. La función logística normalizada es una sigmoide descrita por la ecuación 1.7 [34]

$$\sigma = \frac{1}{1 + e^{-t}} \quad (1.7)$$

En donde t es un valor dado por la combinación lineal de las características de las instancias. Se dice que esta función está normalizada, a partir de la función logística general, ya que σ solo toma valores reales entre 0 y 1. Luego de realizar un ajuste con los datos de entrenamiento, se obtienen los parámetros de ajuste que permiten realizar el cálculo de la probabilidad de pertenencia a la clase objetivo para cada instancia de los datos de prueba. Después de calcular la probabilidad de pertenencia, la instancia se clasifica como positiva o perteneciente a dicha clase si la probabilidad $\hat{p} \geq 0.5$, o como perteneciente a la otra clase si $\hat{p} < 0.5$, cabe mencionar que este punto de corte puede ser modificado si se requiere dar más peso a alguna de las clases. Para realizar el entrenamiento del modelo, es necesario utilizar funciones de costo que permitan asignar valores de probabilidad mayores a las instancias pertenecientes a la clase objetivo, y valores de probabilidad menores a las instancias pertenecientes a la otra clase. Este modelo también puede ser utilizado para tareas de regresión.

1.4.5. Máquinas de soporte vectorial

Otro modelo de aprendizaje autónomo muy útil en las tareas de clasificación es el denominado máquinas de soporte vectorial o SVM (Support Vector Machine), por sus siglas en inglés.

Este conjunto de algoritmos se utilizan para tareas de clasificación y regresión y se fundamentan en separar gráficamente las diferentes clases de un conjunto de datos. Para lograr separar los conjuntos de datos, se busca encontrar un hiperplano que maximice la separación entre los puntos de ambos conjuntos, como se aprecia en la figura 1.9.

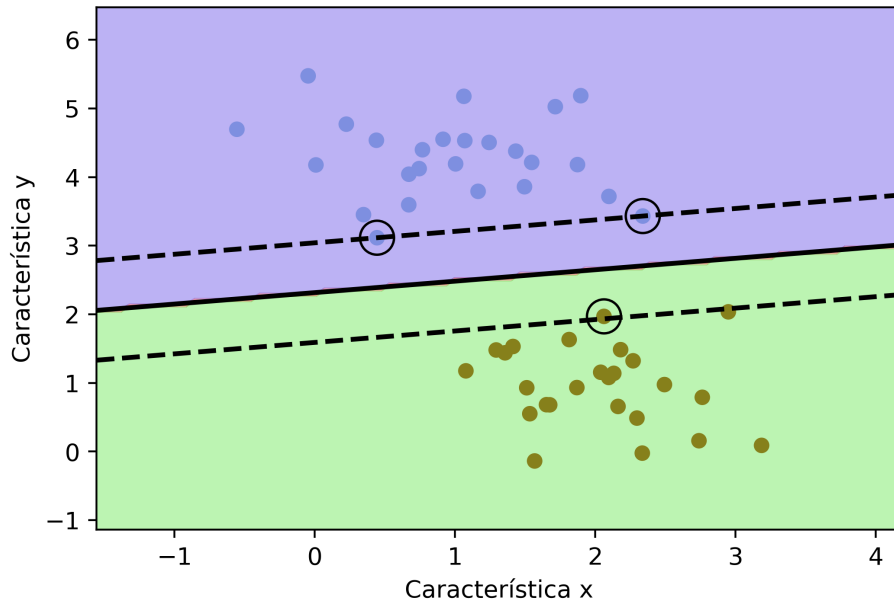


Figura 1.9: Representación gráfica de la clasificación a través de un algoritmo de SVM.

Se observa que el hiperplano consta de un vector central y dos vectores laterales, estos se utilizan para obtener un margen que permita delimitar correctamente las dos clases de datos, ya que al añadir nuevas instancias estas pueden ser clasificadas erróneamente si no se encuentra el vector que separe de mejor manera los datos. Los vectores laterales generalmente se definen por los datos más cercanos al vector central y estos se conocen como vectores de soporte. El hiperplano se define de manera tal que los márgenes estén optimizados y que sean iguales en ambos lados. Matemáticamente, el algoritmo tiene como objetivo minimizar la suma de los cuadrados de las desviaciones de los vectores de soporte de la frontera de decisión, sujeto a la restricción de que cada vector de soporte esté correctamente clasificado [34]. Para esto primero se aplica una técnica llamada solución dual que permite transformar el problema original en uno más simple de resolver, luego se utilizan técnicas de optimización como los multiplicadores de Lagrange para hallar la solución óptima.

Es importante mencionar que en el ejemplo mostrado, los datos son linealmente separables, es decir, que es posible separarlos con un hiperplano lineal, sin embargo, en muchos problemas los datos no se comportan de esta manera y es necesario realizar modificaciones al algoritmo para lograr dividir los datos correctamente. Una de las soluciones a este problema es añadir un parámetro C al algoritmo, el cual permite modificar el hiperplano de manera tal que se ajuste a los datos, sin embargo, este ajuste también es lineal. Para los datos que no pueden ser separados linealmente, generalmente se tienen dos opciones: aplicar una función de transformación polinomial o aplicar un kernel polinomial. En el primero de los casos simplemente se aplica una función de tipo $P = x^n$ a los datos, con la finalidad de hacerlos linealmente separables.

La aplicación de los kernel polinomiales en cambio, consiste en separar los datos en espacios de características con mayor dimensión, en donde los datos pueden ser linealmente separables, esto da como resultado un hiperplano no lineal en el espacio original de características [34]. También existen otros tipos de kernel como los lineales y sigmoideos.

1.4.6. K vecino más cercano

El modelo de los k vecinos más cercanos o KNN (K-Nearest Neighbor), por sus siglas en inglés, permite clasificar una instancia a través de la asignación de pertenencia a la clase del vecino o de los vecinos más cercanos del dato en cuestión. Para realizar esto, primero almacena todas las instancias de entrenamiento y, para predecir la clase de una nueva instancia, calcula la distancia a los k vecinos más cercanos en el conjunto de entrenamiento y toma la clase más común de estos k vecinos como la predicción para la nueva instancia [36], como se aprecia en la figura 1.10.

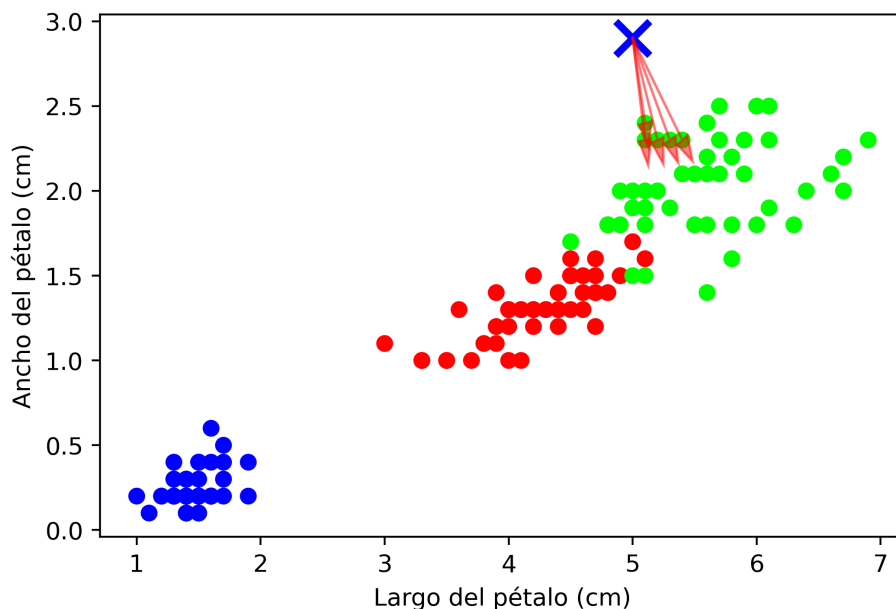


Figura 1.10: Representación gráfica de un ejemplo del cálculo de los k vecinos más cercanos para predecir la clase de una instancia.

El cálculo de los k vecinos más cercanos, normalmente se realiza midiendo la distancia euclidiana, pero existen otros métodos como el cálculo de la distancia Manhattan, la elección del método para calcular las distancias estará sujeta a las características del problema. Otro factor a tomar en cuenta es el número de k vecinos que se van a considerar para la clasificación, ya que si se elige un número pequeño, la predicción se vuelve muy sensible al ruido ocasionado por los datos que no pertenecen a la clase correcta, mientras que al elegir una k muy grande, puede que más instancias de la clase incorrecta queden dentro de la distancia euclidiana, reduciendo la precisión de la clasificación [36]. Para elegir un número de k vecinos adecuado normalmente se emplean métodos de validación cruzada, el cual es un algoritmo que realiza una operación específica variando alguno de los parámetros que afectan la operación y elige el parámetro con el mejor resultado para realizar la operación. El modelo KNN permite obtener buenos resultados

en cuanto a predicciones, superando a otros métodos más sofisticados si se utiliza adecuadamente, además, es fácil de modificar para resolver problemas complejos con más clases. Sin embargo, se debe considerar que puede requerir mayor tiempo de cómputo debido a que para cada instancia se calculan las distancias con todos los datos de entrenamiento [36].

1.4.7. Redes neuronales

Las redes neuronales artificiales o simplemente redes neuronales son un grupo de algoritmos originalmente basados en las estructuras neuronales cerebrales, las cuales se utilizan como herramientas de regresión y clasificación bimodal o multimodal. Estos algoritmos surgieron de modelos más simples en los cuales se tienen estructuras de entrada y estructuras de salida, ambas conectadas por redes. Estas estructuras pueden tener varias entradas y varias salidas conectadas entre sí, como se observa en la figura 1.11.

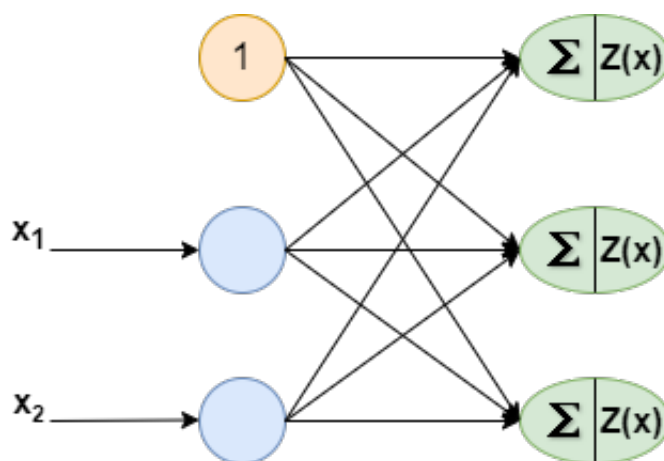


Figura 1.11: Representación gráfica de una red neuronal simple.

Este arreglo permite que a partir de ciertas entradas, que en un problema serían las características, se obtengan salidas binarias que nos indiquen una operación o clasificación, cabe mencionar que las entradas pueden ser valores numéricos que no se encuentren entre 0 y 1, ya que las redes mediante las que se conectan con las salidas tienen un peso w diferente para cada conexión, el cual permite que los valores de entrada se sumen en cada salida y, al aplicar una función de activación, la salida dé un valor entre 0 y 1. A este tipo de redes neuronales que aplican pesos y sumatorias a sus conexiones de entrada y aplican funciones de activación en las salidas se le conoce como Perceptrones y matemáticamente el proceso de obtención de sus valores de salida se expresa de la siguiente manera para una red con varias entradas [34]:

$$h_{\mathbf{W},\mathbf{b}}(\mathbf{X}) = \phi(\mathbf{X}\mathbf{W} + \mathbf{b}) \quad (1.8)$$

En donde \mathbf{X} representa la matriz de características de entrada para todas las instancias de entrenamiento, \mathbf{W} es la matriz con los pesos de todas las conexiones, ϕ es la función de activación y \mathbf{b} es un vector que contiene los pesos de las conexiones de cierto tipo de neuronas llamadas neuronas parciales, las cuales siempre tienen un valor de 1. Para la función de activación normalmente se emplea una función logística de la forma de la ecuación (1.7), la cual permite obtener como salida una probabilidad de pertenencia a una clase objetivo, sin embargo,

también se pueden emplear otras funciones de activación como la función tangente hiperbólica o la función escalón cuando se desean salidas binarias. Para entrenar un algoritmo de red neuronal tipo Perceptrón simple, el principio consiste en reforzar las conexiones que dan como resultado una predicción correcta y modificar las conexiones que propician una predicción errónea. Para esto, el algoritmo utiliza los datos de entrenamiento de una instancia de la cual se conoce previamente su clase y realiza una predicción con pesos $w_{i,j}$ (i-ésima entrada y j-ésima salida) aleatorios para cada conexión, si la predicción es errónea el algoritmo baja los pesos de las conexiones que influyen más en dicha predicción, hasta obtener una predicción correcta. Esto matemáticamente se puede expresar de la siguiente manera [34]

$$w_{i,j}^{(\text{siguiente iteración})} = w_{i,j} + \eta (y_j - \hat{y}_j) x_i \quad (1.9)$$

En donde x_i es el i-ésimo valor de entrada de la instancia de entrenamiento actual, \hat{y}_j es la salida de la j-ésima neurona de salida de la instancia de entrenamiento actual, y_j es la salida deseada de esta misma neurona de salida y η es la tasa de aprendizaje. Cabe mencionar que este proceso se aplica instancia por instancia con los datos de entrenamiento. Las redes neuronales tipo perceptrones simples son capaces de resolver problemas sencillos, pero para problemas más complejos no son efectivas, debido a esto se desarrollaron algoritmos más complejos que utilizan múltiples capas de perceptrones denominados Perceptrones Multi-Capa o MLPs, por sus siglas en inglés. Este tipo de algoritmos, además de contener las capas correspondientes a las neuronas de entrada y salida, contienen más capas de neuronas intermedias denominadas capas ocultas como se ilustra en la figura 1.12, las cuales en conjunto con un algoritmo de retropropagación hacen posible obtener muy buenos resultados en las predicciones de problemas con patrones más complejos.

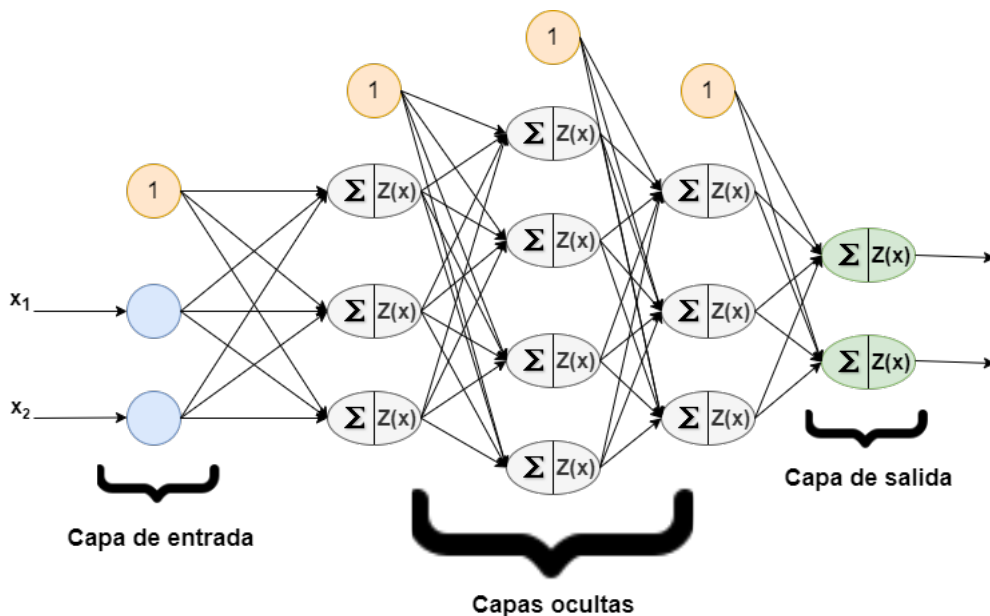


Figura 1.12: Representación gráfica de una red neuronal tipo Perceptrones Multi-Capa.

El algoritmo de retropropagación de manera general permite realizar correcciones en los pesos de las conexiones de las neuronas de manera que, al realizar la predicción para una instancia

de los datos de entrenamiento, se evalúa el error en la predicción y el algoritmo va en reversa a través de las capas neuronales para medir la contribución errónea de cada conexión, finalmente el algoritmo modifica los pesos de dichas conexiones para reducir el error. Este proceso se realiza para cada instancia de entrenamiento y cabe mencionar que, para que el algoritmo llegue a la solución óptima, los pesos de las conexiones deben estar distribuidos aleatoriamente, esto para evitar que la contribución de las conexiones sea parecida y por lo tanto que las correcciones sean parecidas, ocasionando una predicción imprecisa. Cuando un algoritmo tipo MLP contiene muchas capas ocultas se le denomina red neuronal densa o profunda (DNN, por sus siglas en inglés), estas son una de las herramientas utilizadas para el aprendizaje profundo o Deep Learning [34].

1.4.8. Aplicación de modelos de aprendizaje autónomo en medicina

Se han realizado numerosos estudios aplicando algoritmos de aprendizaje autónomo ya que potencialmente pueden ser utilizados para identificar fenotipos tumorales [37], también se pueden aplicar para dar seguimiento a un tratamiento y observar el progreso en el control tumoral [38]. Además, también es posible aplicar estas herramientas para identificar o predecir el diagnóstico de otras enfermedades en las que se requiere un diagnóstico por imagen, por ejemplo, para la Enfermedad Alzheimer se realizan con frecuencia estudios de imagen que permiten observar las regiones cerebrales atrofiadas debido a la enfermedad y esto permite realizar estudios en los que se extraen características radiómicas de dichas zonas o regiones de interés y entrenar modelos de aprendizaje autónomo con los datos obtenidos. Por ejemplo, en un estudio se analizaron imágenes $[^{18}\text{F}]$ FDG-PET de pacientes con sospecha de EA, mediante la extracción de características y entrenamiento utilizando un modelo de redes neuronales convolucionales [7], en el cual se obtuvieron buenos resultados. También se han realizado estudios en los que se emplean las redes neuronales convolucionales para obtener las características radiómicas y posteriormente entrenar un modelo de SVM para clasificar pacientes con riesgo de padecer EA y pacientes neurológicamente sanos de la base de datos de la Iniciativa de Neuroimagen de Enfermedad de Alzheimer (ADNI, por sus siglas en inglés) [39]. En otros estudios, las características radiómicas se midieron directamente de las regiones de interés previamente segmentadas gracias a un preprocesamiento de las imágenes y al uso de programas de segmentación, utilizando una muestra de 43 pacientes (con sospecha de EA y controles), para después entrenar modelo de aprendizaje supervisado (análisis discriminante) con el que se hicieron las predicciones [40]. Este mismo procedimiento se ha realizado en otros estudios en donde además se emplean métodos de análisis estadístico para obtener las características radiómicas más relevantes y eliminar las que no tienen importancia en el entrenamiento del modelo [41]. Además de los estudios que analizan la captación de glucosa para el diagnóstico de la EA, también se han realizado análisis de imágenes PET para estudios realizados con biomarcadores que posiblemente pueden describir el proceso de evolución de la enfermedad desde sus etapas tempranas [7], e incluso en algunos estudios se ha demostrado que es posible predecir los casos en los que un deterioro cognitivo evolucionará a la EA, empleando imágenes $[^{18}\text{F}]$ FDG-PET en conjunto modelos de aprendizaje autónomo y radiómica [8].

En el presente trabajo se utilizó una metodología diferente para realizar la medición de características radiómicas, de manera que estas fueron medidas en regiones cerebrales definidas por un atlas, estas regiones representan partes específicas del cerebro, en vez de combinar varias

partes del cerebro como en trabajos anteriores. Además, para las predicciones se utilizaron varios algoritmos de aprendizaje autónomo, comparando sus resultados. Para el entrenamiento de estos algoritmos se utilizaron las características medidas en todas las regiones, a diferencia de trabajos anteriores en los que el análisis se hace región por región. Otra diferencia importante con trabajos anteriores es que se realizaron diferentes casos de estudio y en uno de estos se incluyeron pacientes con enfermedades neurológicas, esto es importante para que los algoritmos de predicción no solo sean sensibles, sino también específicos. Además, a pesar de utilizar estudios de pacientes de un solo centro de imagen PET, la muestra es ligeramente superior a la utilizada en trabajos anteriores en donde solo usaron estudios de un centro.

1.5. Justificación

El Alzheimer afecta a un porcentaje importante de la población mayor de 65 años, siendo un problema de salud pública ya que es irreversible y progresiva. Se ha observado que las etapas tempranas de la enfermedad (deterioro cognitivo) pueden presentarse varios años antes y evolucionar a Alzheimer en algún momento, también se sabe que no todos los pacientes que presentan deterioro cognitivo llegan a tener Alzheimer, sin embargo, los mecanismos que definen si un paciente con deterioro cognitivo evolucionará a Alzheimer aún son inciertos. Uno de los problemas que se presentan con la enfermedad es su detección, ya que en la actualidad la prueba más confiable de un diagnóstico de Alzheimer es el análisis histopatológico postmortem. Las técnicas de imagen funcional, son una herramienta de apoyo en el diagnóstico de la enfermedad de Alzheimer, en especial las de medicina nuclear que permiten obtener información cualitativa del comportamiento de moléculas biológicas importantes durante la enfermedad. La PET es una de las técnicas más atractivas debido a la capacidad de observar cambios metabólicos, los cuales anteceden a los cambios anatómicos. Los estudios PET con el radiofármaco $[^{18}\text{F}]\text{FDG}$ son los más utilizados y en los estudios de demencias cobran relevancia debido a la utilidad de observar cambios en el metabolismo de glucosa, asociado con la densidad sináptica e integridad neuronal. Los factores que propician la presencia de la EA, se pueden presentar como patrones metabólicos específicos en estudios PET que deben ser evaluados y medidos cuantitativamente, lo cual conlleva a la necesidad de herramientas que permitan analizar grandes cantidades de datos para realizar estimaciones o predicciones.

Los algoritmos de Machine Learning (aprendizaje autónomo) supervisados pueden llegar a ser una herramienta importante para esta tarea, ya que son capaces de evaluar grupos grandes de datos, identificar un patrón y predecir un resultado para otro grupo de prueba. Se han realizado múltiples estudios con algoritmos de aprendizaje autónomo en imagen médica para predecir diagnósticos de enfermedades con resultados excelentes. Para apoyar en el diagnóstico de la Enfermedad de Alzheimer con imagen médica, específicamente con PET, se han realizado pocos estudios utilizando estos algoritmos, pero se han obtenido resultados favorables y prometedores. Es necesario realizar más estudios al respecto, además, aplicar diferentes herramientas de extracción de datos de las imágenes como la radiómica puede mejorar el rendimiento de los algoritmos, por lo que queda justificada la realización de esta investigación.

1.6. Hipótesis

Es posible mejorar un diagnóstico de EA a partir de un grupo de imágenes $[^{18}\text{F}]$ FDG-PET extrayendo sus características radiómicas, identificando su relación con la EA y empleando modelos de aprendizaje supervisado.

1.7. Objetivos

Objetivo principal

- Evaluar algoritmos de aprendizaje autónomo para realizar predicciones del diagnóstico de pacientes con sospecha de Enfermedad de Alzheimer, a partir de características de intensidad y textura obtenidas de regiones cerebrales de los pacientes.

Objetivos particulares

- Obtener características de intensidad y textura de diferentes regiones cerebrales en estudios $[^{18}\text{F}]$ FDG-PET de pacientes con sospecha de Enfermedad de Alzheimer.
- Correlacionar características de intensidad y textura de estudios $[^{18}\text{F}]$ FDG-PET con el diagnóstico dado por médicos radiólogos y nucleares.
- Obtener mapas de probabilidad (t) de estudios $[^{18}\text{F}]$ FDG-PET para clasificarlos como pacientes con o sin Enfermedad de Alzheimer.
- Entrenar diferentes algoritmos de aprendizaje autónomo a partir de características de intensidad y textura obtenidas de regiones cerebrales de los pacientes.
- Evaluar el desempeño de los algoritmos de aprendizaje autónomo en la clasificación de pacientes mediante diferentes métricas.

Capítulo 2

Metodología

En este trabajo se realizó un estudio retrospectivo con imágenes $[^{18}\text{F}]$ FDG-PET, enfocado en pacientes con sospecha de padecimiento de la Enfermedad de Alzheimer. A partir de la extracción de características de intensidad y textura en los volúmenes que conforman el cerebro, se buscaron características que presentan una mayor correlación con el diagnóstico de los pacientes y con una clasificación obtenida mediante un software especializado en análisis de la EA. El proceso para realizar este trabajo consta de las siguientes etapas:

1. Obtención de las imágenes PET
2. Obtención de mapas de t valores y clasificación
3. Preprocesamiento de imágenes PET
4. Segmentación de los volúmenes de interés
5. Adquisición de características de radiómica
6. Análisis estadístico
7. Entrenamiento de los algoritmos
8. Pruebas y predicciones
9. Elección de características importantes y pruebas finales

De manera general, el flujo de trabajo que se siguió se presenta en la figura 2.1 y cada etapa se describe a detalle en las siguientes secciones.

2.1. Obtención de las imágenes PET

En este trabajo se analizaron imágenes PET corregidas por atenuación de 75 sujetos a los cuales se les realizó un estudio con $[^{18}\text{F}]$ FDG en la Unidad PET/CT de la Facultad de Medicina de la UNAM, cabe mencionar que estos pacientes provienen de otros centros de salud y llegaron referidos a la unidad con diferentes diagnósticos de envío. Los pacientes fueron seleccionados tomando en cuenta los criterios de inclusión y exclusión descritos en la figura 2.2.

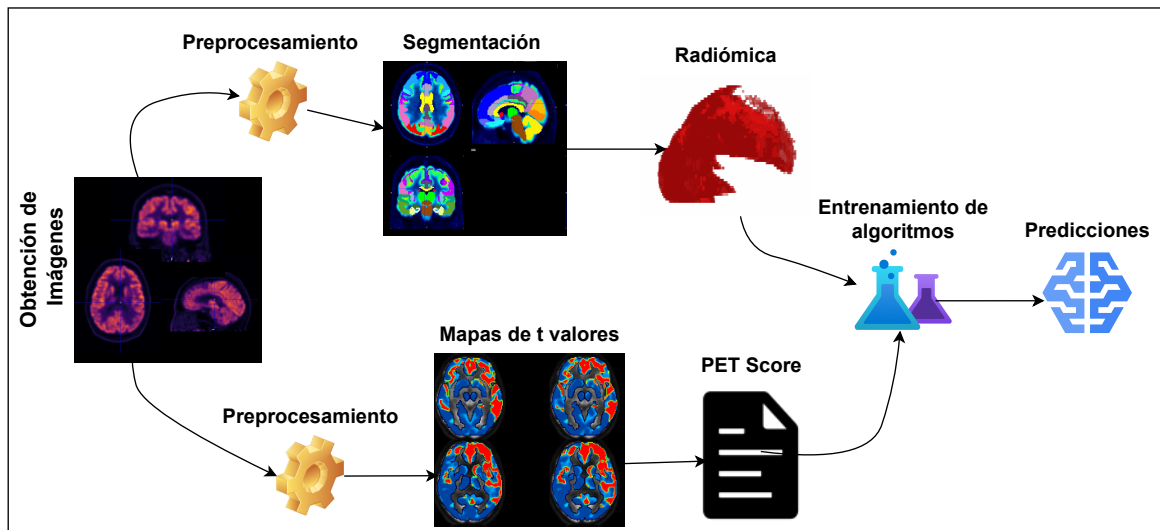


Figura 2.1: Etapas y flujo de trabajo a seguir.

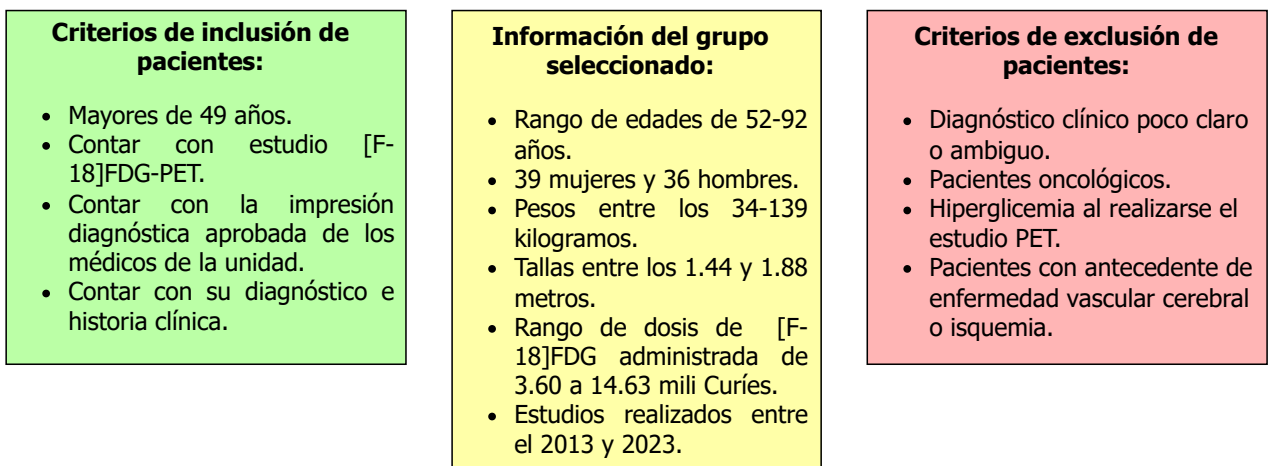


Figura 2.2: Información general de los pacientes y criterios de inclusión y exclusión.

También se proporcionaron los datos de cada paciente con los que cuenta la unidad para conocer su información demográfica, su diagnóstico clínico (de envío) y la impresión diagnóstica de salida dada por los médicos radiólogos y nucleares de la unidad, esta información formará parte de los datos utilizados para entrenar a los algoritmos de aprendizaje autónomo, como se explica más adelante. La información demográfica proporcionada consta de la edad, género, peso y talla de cada paciente. Para la reconstrucción de las imágenes PET se empleó el algoritmo de reconstrucción OSEM 2D y fueron adquiridas en un equipo Siemens Biograph por el personal de la unidad.

Los 75 pacientes constan de tres grupos: 25 pacientes con sospecha de EA en su historia clínica, 25 pacientes con alguna enfermedad neurológica (Enfermedad de Parkinson, Epilepsia, Esquizofrenia, etc) y 25 pacientes controles. Los pacientes controles no presentaron alguna enfermedad neurológica en su diagnóstico clínico ni en la impresión diagnóstica de los médicos de la unidad y constan principalmente de pacientes que formaron parte de algún protocolo de

estudio, pacientes con pérdida de peso (sin alteración metabólica de glucosa encontrada), pacientes con enfisema pulmonar y pacientes con alguna formación benigna no catalogada como cáncer, estos también fueron elegidos tomando en cuenta los criterios de inclusión y exclusión descritos en la figura 2.2.

2.2. Obtención de mapas de t valores y clasificación

Se obtuvieron mapas de t valores utilizando como entrada las mismas imágenes PET de cada paciente, esto con la finalidad de calcular un valor numérico que permitió clasificar a los pacientes como normales o como posible Alzheimer. Este proceso se realizó mediante el software PMOD [42] en el módulo *Alzheimer Discrimination Tool*, iniciando con una normalización de las imágenes y aplicándoles un filtro Gaussiano de 12 mm para reducir el efecto de los bordes [43]. Posteriormente, se dividió el valor de intensidad de cada voxel de las imágenes entre un valor medio de voxel, el cual se obtuvo para cada paciente promediando los valores de intensidad dentro de una máscara de voxels que preservan su actividad con respecto a la EA, es decir, son regiones de referencia que no presentan cambios en su actividad en presencia de la EA. Luego, se estimó la actividad esperada en cada voxel de la imagen de acuerdo con la edad del paciente, utilizando parámetros obtenidos en una regresión de edad. Se compararon los valores de la actividad que presentan los voxels de las imágenes de los pacientes, con los valores de actividad esperada en cada voxel de acuerdo con la edad de los pacientes, esta comparación se realizó calculando valores t de cada voxel, los cuales representan la probabilidad de que la actividad del voxel medido, sea diferente a la actividad del voxel estimado, esto quiere decir que mientras mayor sea el valor t del voxel, es más probable que este sea diferente del valor esperado de acuerdo a su edad. Los valores t para cada voxel se calcularon mediante la siguiente expresión

$$t = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \quad (2.1)$$

en donde \bar{X} es el valor medio del voxel medido de la imagen del paciente, μ es el valor esperado de acuerdo con su edad, σ es la desviación estándar de los voxels y n el número de voxels [44]. Luego se seleccionaron los voxels de las regiones que muestran una correlación con la EA de acuerdo con el Mini Mental Status Examination (MMSE) mediante la aplicación de una máscara, se sumaron los t valores correspondientes a dichos voxels y se compararon con un valor umbral establecido de 11089, el cual fue calculado con un intervalo de confianza del 95% en un estudio previo y representa el límite que se puede obtener en la suma de los t valores en pacientes normales [43]. La comparación con el valor umbral se realizó a través del cálculo de un valor numérico denominado puntaje PET (PET Score), el cual se obtiene mediante la siguiente expresión

$$punteje\ PET = \log_2 \left(\frac{Suma(t_{valores})}{11089} + 1 \right) \quad (2.2)$$

si el valor del puntaje PET es mayor que 1, el paciente es clasificado como anormal respecto a la EA, es decir, como probable Alzheimer [45]. Este valor también es un indicativo indirecto de la severidad de la enfermedad. Cabe mencionar que los mapas de t valores obtenidos de este análisis, representan los voxels de las regiones correlacionadas con la EA en escala de valores t , fusionados con la imagen normalizada de cada paciente. En la figura 2.3 se observa un resumen del proceso para obtener los mapas de t valores y el puntaje PET (PET Score).

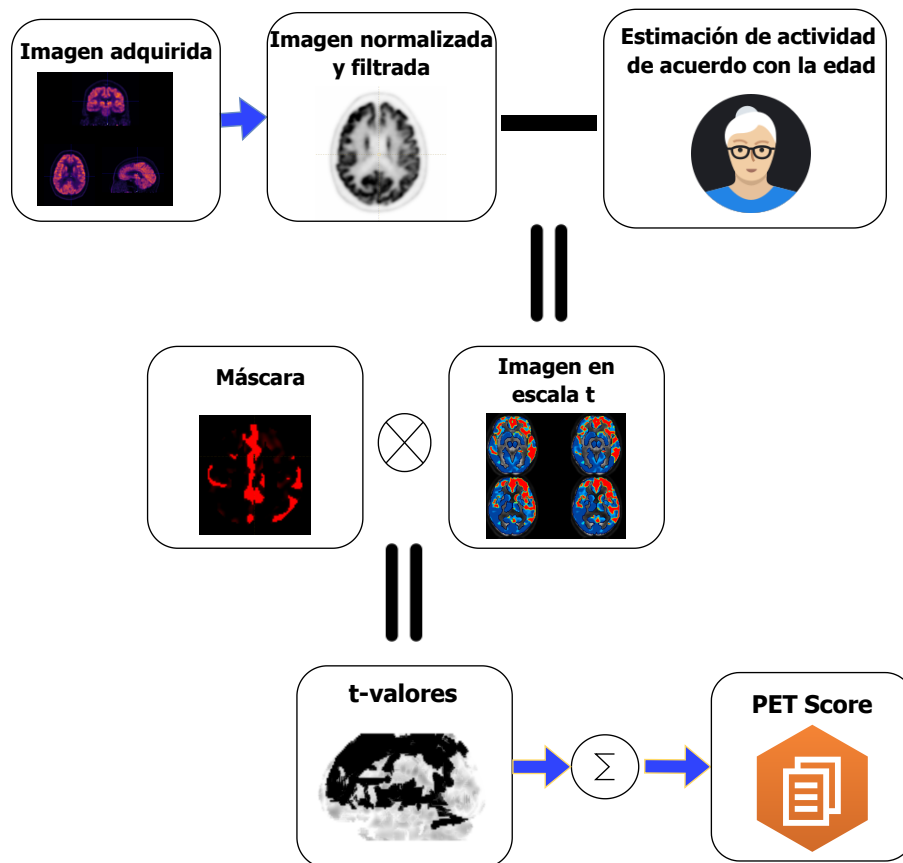


Figura 2.3: Ilustración del procedimiento para obtener el puntaje PET (PETScore) en el PMOD Alzheimer Discrimination Tool.

2.3. Preprocesamiento de imágenes PET

Además del preprocesamiento realizado en el procedimiento para obtener los mapas de t valores, también se realizó un preprocesamiento de las imágenes PET originales para realizar la adquisición de características radiómicas. Las imágenes se normalizaron espacialmente, utilizando un muestreo de $2 \times 2 \times 2 \text{ mm}^3$, al tamaño del atlas cerebral HAMERS-N30R83 [46] mediante el software de análisis de imagen PMOD. La normalización espacial consiste en transformar el tamaño de la imagen original al tamaño de una imagen de referencia, en este caso del atlas. Este proceso se debe llevar a cabo para que todas las mediciones se realicen en las mismas regiones y no existan variaciones ocasionadas por medir en lugares diferentes en la imagen. Cada grupo de imágenes correspondientes a los cortes por paciente fueron normalizadas y se exportaron también como grupo en su formato original, esto es importante para realizar el cálculo de los valores SUV en la imagen.

2.4. Segmentación de los volúmenes de interés

Posterior a la normalización de las imágenes al espacio del atlas HAMERS-N30R83, se generaron los volúmenes de interés utilizando como base este mismo atlas en el software PMOD,

tomando las siguientes estructuras cerebrales: lóbulo temporal, amígdalas, giro fusiforme, hipocampo, giro temporal, giro parahipocampal, ínsula, cíngulo, lóbulo frontal, lóbulo parietal, lóbulo occipital, cerebelo, tronco encefálico, cuerpo calloso, tálamo, pálido, putamen, núcleo accumbens, núcleo caudado y materia blanca. Las estructuras correspondientes a estas regiones fueron desplegadas sobre las imágenes PET normalizadas para medir las características radiómicas, como se ilustra en la figura 2.4.

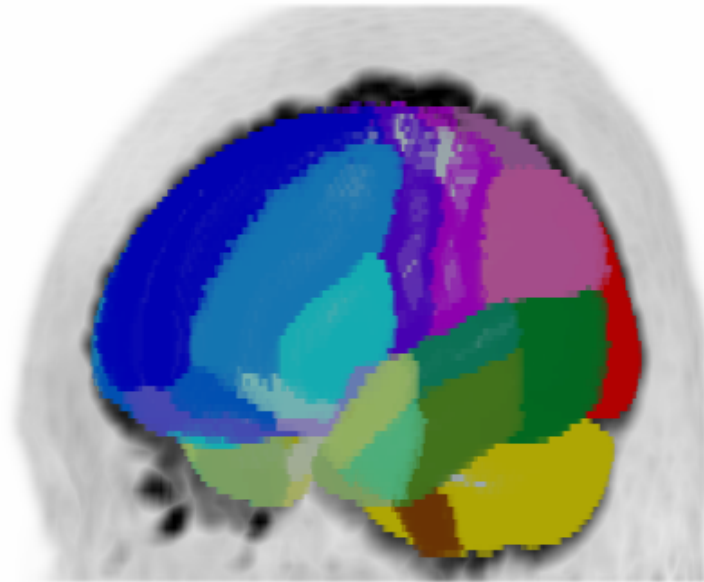


Figura 2.4: Visualización de los volúmenes de interés desplegados sobre una de las imágenes normalizadas.

2.5. Adquisición de las características de radiómica

Las características de radiómica se calcularon para cada volumen de interés en las imágenes normalizadas de cada paciente mediante el software LIFEx, el cual es uno de los programas de radiómica reconocidos por la iniciativa de estandarización de biomarcadores de imagen o IBSI [47], por sus siglas en inglés. Las mediciones se tomaron utilizando un muestreo de tamaño de voxel de $2 \times 2 \times 2 \text{ mm}^3$, realizando una discretización de la intensidad absoluta y con un número bins de 64, estos parámetros son los recomendados en la documentación del software [33]. Esta discretización limita el número de niveles de gris de la imagen a un número fijo, en este caso 64, por lo que los niveles de gris intermedios de la escala original toman un valor dentro de la nueva escala, esto se puede entender como redondear o truncar los valores de los niveles de gris. Se calcularon características de primer y segundo orden, las cuales se describieron previamente y el fundamento matemático se puede encontrar en los anexos de este trabajo. Las características medidas de todos los volúmenes de interés para cada paciente se exportaron y se ordenaron en un único archivo, descartando las características redundantes y también las características de tamaño y forma ya que todos los volúmenes de interés son los mismos. Cabe mencionar que las imágenes se encontraban en escala SUV al realizar las mediciones.

2.6. Análisis estadístico

Como se mencionó, los 75 estudios constan de 25 pacientes con sospecha de Enfermedad de Alzheimer, 25 con alguna enfermedad neurológica y 25 controles. Para realizar el análisis estadístico y las pruebas con los algoritmos de aprendizaje supervisado se formaron dos grupos, uno con los 25 pacientes con sospecha de EA y con los 25 pacientes con alguna enfermedad neurológica, y el otro con los mismos 25 pacientes con sospecha de EA y con los 25 controles. En ambos grupos la clase objetivo del estudio fue el diagnóstico clínico, de manera que los pacientes con sospecha de EA fueron clasificados como “Alz” y los pacientes con enfermedad neurológica y controles fueron clasificados como “PC”. Adicionalmente, para el grupo de pacientes con sospecha de EA y controles, se realizó el mismo estudio utilizando como clase objetivo la clasificación dada por la herramienta del software PMOD, de manera que se realizaron tres estudios diferentes como se ilustra en la figura 2.5, utilizando en cada uno 50 pacientes.

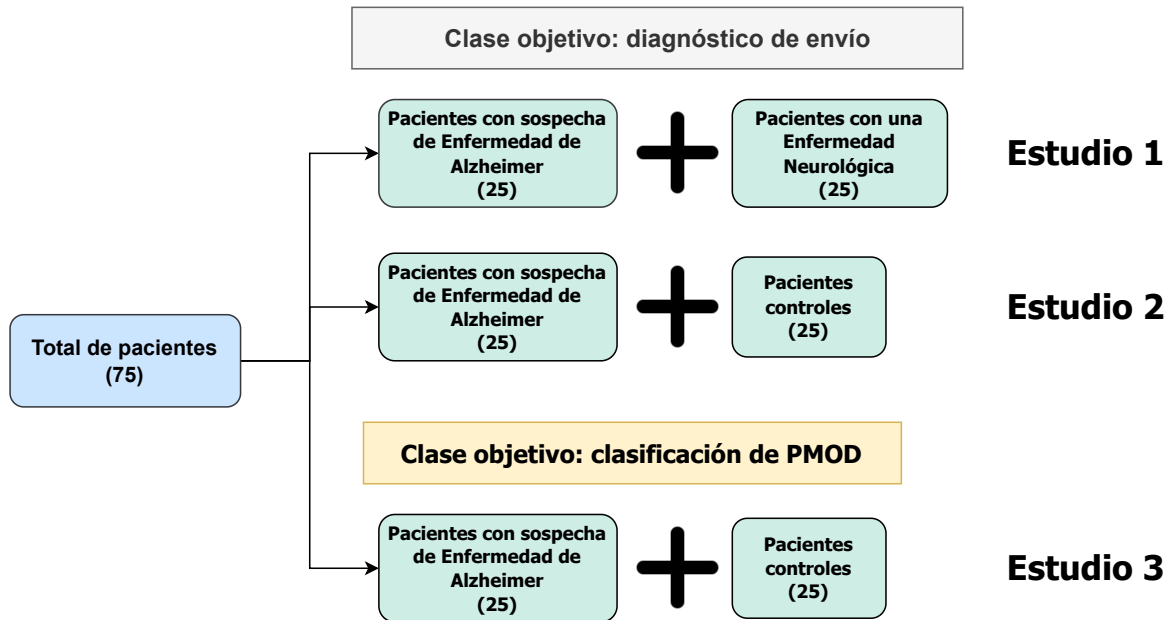


Figura 2.5: Diagrama ilustrativo de la división de pacientes para los diferentes estudios realizados.

Se realizó un análisis para estimar la correlación que se tiene entre las características de radiómica medidas para cada región cerebral, la información demográfica de los pacientes (edad, peso y talla), el puntaje PET y clasificación obtenidos del software PMOD y el diagnóstico clínico de los pacientes. Para esto se realizó una prueba de correlación de Pearson, la cual consiste en estimar la correlación lineal entre dos variables a través del método de mínimos cuadrados, de manera que al realizar una regresión lineal entre ambas variables por el método de mínimos cuadrados y calcular una recta de ajuste de tipo $y = a + bx$, es posible estimar la correlación r entre ambas variables mediante la siguiente expresión [44]

$$r^2 = \frac{b^2 \left[\sum x_i^2 - (\sum x_i)^2 / n \right]}{\sum y_i^2 - (\sum y_i)^2 / n} \quad (2.3)$$

En donde x_i e y_i son los valores i -ésimos de cada variable en una muestra de tamaño n y b es la pendiente de la recta de regresión. Este coeficiente es un indicativo de la correlación entre ambas variables, de manera que si $r = 1$ ó $r = -1$, se tiene una correlación perfecta directa o inversa entre ambas variables, respectivamente, y si $r = 0$, significa que no existe correlación alguna. Cabe mencionar que es posible elegir cualquiera de las dos variables como x e y debido a que en el modelo de correlación ambas variables son aleatorias, pero es preferible elegir a la variable sobre la que se desean hacer las predicciones como la variable dependiente de la regresión [44]. En el caso de este trabajo, las variables x e y a comparar representan todas las características empleadas para el entrenamiento de los algoritmos. El análisis de correlación de Pearson se realizó de manera separada para los dos grupos de 50 pacientes descritos anteriormente con el objetivo de observar qué características medidas en cada estructura cerebral en las imágenes PET, están más correlacionadas con el diagnóstico clínico de los pacientes y con el puntaje PET obtenido en el módulo de Alzheimer de PMOD.

En este análisis se utilizó el coeficiente de correlación de Pearson para seleccionar las características que parecen estar más relacionadas con el diagnóstico clínico, sin embargo, este método puede no ser el adecuado para inferir una correlación entre las características y el diagnóstico clínico, ya que en este se tienen únicamente dos valores que son Alzheimer y enfermedad neurológica o controles, dependiendo del caso de estudio, por lo que si asignamos el valor de “1” a los pacientes con Alzheimer y de “0” a los pacientes controles o con enfermedad neurológica, aún así tendríamos una variable con solo dos valores. Por lo tanto, además del coeficiente de correlación de Pearson, se calculó el coeficiente de correlación de Spearman, en el cual también se realizó una regresión de tipo lineal para calcular el coeficiente de correlación, pero en este caso se hizo por rangos en una de las variables, esto permite que el método pueda ser utilizado para variables nominales siempre y cuando se les asigne un valor numérico categórico [48]. A la variable de género también se le asignaron valores numéricos de “1” para el femenino y “0” para el masculino, sin embargo no fue considerada en los análisis de correlación ya que no se puede comparar una variable categórica contra otra categórica, pero si fue incluida en las pruebas utilizando todas las características que se describen más adelante.

Luego de obtener las características con mayores coeficientes de correlación de Spearman con el diagnóstico clínico, fue necesario comprobar que los valores de dichas características en ambos tipos de pacientes son diferentes mediante pruebas no paramétricas, ya que los datos probaron tener una distribución no normal al aplicarles una prueba de Kolmogorov-Smirnov [49] y también al aplicar una prueba de Shapiro-Wilk [50]. Un test no paramétrico que permite obtener una medida acerca de la diferencia entre tres o más grupos de datos es la prueba de Kruskal-Wallis, cuando solo se tienen dos grupos de datos, la prueba se denomina Mann-Whitney U. En este caso se realizó el análisis para comprobar que las características con mayor coeficiente de correlación de Spearman con el diagnóstico clínico son diferentes entre los dos tipos de pacientes en cada caso de estudio. La prueba de Kruskal-Wallis o en este caso de Mann-Whitney U, da como resultado un valor p y un estadístico de prueba, de manera que mientras menor sea el valor p y mayor sea el estadístico de prueba, mayor será la probabilidad de que la distribución de valores de la característica analizada sea diferente para los dos o más grupos de datos. Concretamente el valor p debe ser menor que un valor de referencia, convencionalmente 0.05, para que la diferencia entre ambos grupos de datos sea estadísticamente significativa [48].

2.7. Entrenamiento de los algoritmos

Para realizar el entrenamiento de los algoritmos de aprendizaje autónomo, se empleó el software *Orange*, el cual permite utilizar una interfaz amigable y sencilla para cargar y filtrar datos, realizar análisis estadístico, entrenar diferentes modelos de aprendizaje autónomo supervisado y no supervisado, modificar sus parámetros, realizar pruebas de rendimiento de los modelos y diversos análisis numéricos y gráficos con los resultados. Previamente se ordenaron las características medidas para cada volumen de interés y se renombraron etiquetando cada una con los nombres de los volúmenes, obteniendo un único grupo de características para cada paciente, este grupo de características contiene las medidas obtenidas para cada región, con esto se espera analizar los patrones hipometabólicos de la EA. Además del grupo de características radiómicas por región, para cada paciente se tiene su edad, género, talla y peso, además del puntaje PET y clasificación del software PMOD obtenidos de los mapas de t valores como se explicó previamente. También se encuentra indicado para cada paciente su diagnóstico clínico. Antes de realizar el entrenamiento y pruebas, se dividieron los dos grupos de 50 pacientes con sus características medidas e información demográfica en datos de entrenamiento, tomando el 70% del total de cada grupo, y datos de prueba, tomando el 30% restante. Se seleccionó la columna objetivo para ambos grupos de pacientes, la cual corresponde a la clase objetivo de los estudios. Como se mencionó, para ambos grupos se utilizó el diagnóstico clínico como clase objetivo y adicionalmente se realizó otro estudio para el grupo de pacientes con sospecha de EA y controles en el cual se tomó la clasificación dada por PMOD como clase objetivo. La selección de pacientes de entrenamiento y prueba se realizó conservando la proporción entre los pacientes con sospecha de EA, etiquetados con las siglas “Alz”, y los pacientes con enfermedad neurológica y controles, etiquetados con las siglas “PC”. Se inició con el entrenamiento de algunos modelos de aprendizaje autónomo supervisado, concretamente Decision Tree, Random Forest, Regresión Logística, Red Neuronal, SVM y KNN, estos realizaron un proceso de “aprendizaje” mediante diferentes algoritmos matemáticos con la finalidad de hacer predicciones acerca de la clasificación de los pacientes. El fundamento matemático del proceso de aprendizaje de cada uno se explicó en el capítulo 1.

2.8. Pruebas y predicciones

Para realizar las primeras pruebas con los algoritmos se eligió el método de muestreo de los datos de entrenamiento, este muestreo consiste en la división del total de los datos de entrenamiento para realizar las primeras pruebas de validación de los algoritmos, si bien recordamos que los datos ya fueron muestreados como datos de entrenamiento y pruebas, los datos de entrenamiento fueron muestreados nuevamente para realizar estas pruebas de rendimiento y verificar la precisión de los modelos en los diferentes casos de estudio. Para esto se eligió la validación cruzada como método de muestreo, este método consiste en realizar diferentes muestreos de los datos y elegir el muestreo con el que se obtienen mejores resultados de rendimiento. En la figura 2.6 se presenta la lógica que se sigue con este algoritmo de muestreo o división de datos.

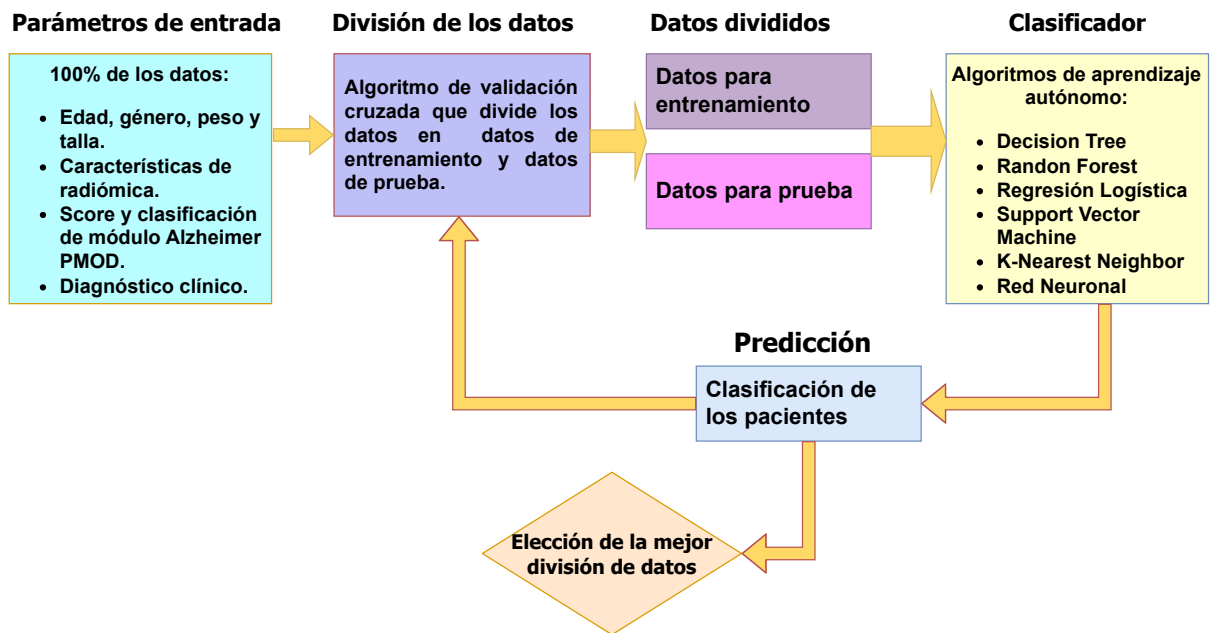


Figura 2.6: Diagrama de la lógica que se sigue al emplear la validación cruzada para dividir los datos de entrada.

También podemos observar cuales son los datos de entrada utilizados en todos los casos de estudio. Una vez elegido el método de muestreo, se realizó el entrenamiento de los diferentes algoritmos y las pruebas correspondientes, obteniendo la predicción del diagnóstico de cada paciente. Se comparó el rendimiento de los modelos mediante el análisis ROC (Receiver Operating Characteristic), el cual evalúa de manera gráfica la tasa de verdaderos positivos (sensibilidad) contra la tasa de falsos positivos ($1 - \text{especificidad}$) predichos por el modelo, obteniendo curvas denominadas como curvas ROC. El área bajo la curva ROC (AUC, por sus siglas en inglés) de un modelo es una medida de su efectividad para predecir el comportamiento de una variable, en este trabajo se eligieron los modelos que obtuvieron una AUC de más de 0.7 en la prueba con validación cruzada, debido a que se considera este valor como una referencia para poder afirmar que un modelo de clasificación es aceptable [51].

Las pruebas anteriormente descritas, solo se realizaron utilizando el archivo con los datos de entrenamiento para verificar la efectividad de los diferentes algoritmos, sin embargo, las pruebas que realmente son válidas para conocer la precisión en la predicción de un algoritmo, son las que se realizan con grupos de datos de prueba que no se encuentran en una misma base, en este caso, en archivos separados. Para esto se emplearon los datos del archivo de entrenamiento, que contiene el 70% del total de los pacientes, para entrenar a los algoritmos y con el archivo que contiene el 30% restante de pacientes se realizaron las predicciones. Esto se hizo para los tres casos de estudio descritos anteriormente. En la figura 2.7 se observa la lógica que se siguió para obtener estas predicciones, además de las características que se utilizaron como parámetros de entrada para entrenar a los algoritmos y predecir su clasificación.

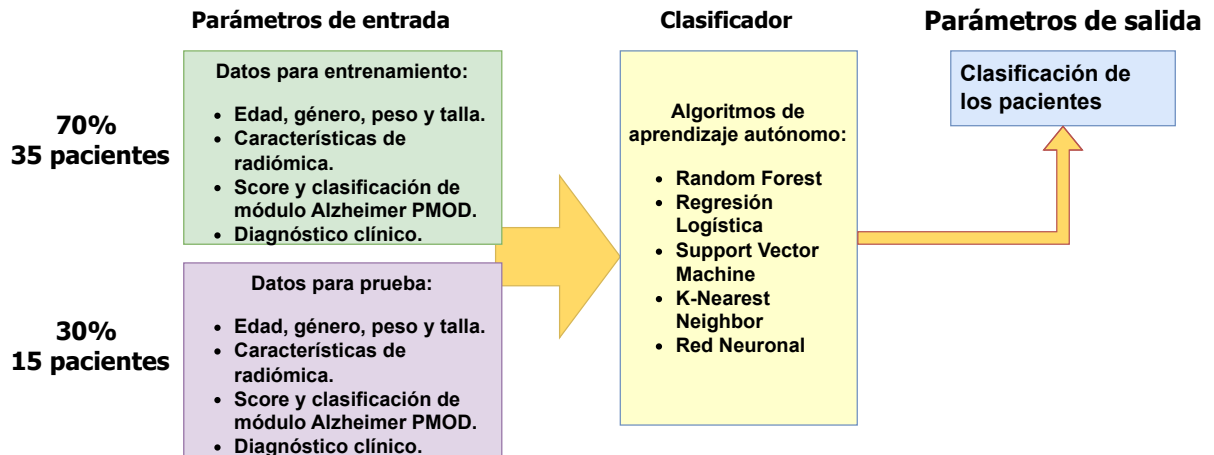


Figura 2.7: Diagrama de la lógica que se siguió al dividir a los pacientes para entrenamiento y prueba.

Es importante verificar que las características de entrada y las clases objetivo de los datos de prueba sean los mismos que los seleccionados en los datos de entrenamiento, ya que de esta información aprenden los algoritmos e identifican la clase de cada paciente de prueba. Cabe mencionar que se normalizaron los datos antes de ser utilizados con los algoritmos de regresión lineal y red neuronal, de manera que la media de la distribución se fijó en cero y la varianza $\sigma^2 = 1$. Para los algoritmos de SVM y KNN, se realizó un análisis de componente principal (PCA, por sus siglas en inglés) [34] utilizando 10 componentes.

2.9. Elección de características importantes y pruebas finales

Finalmente, se realizó la elección de las mejores características con la finalidad de depurar el total de las características que se tienen, eliminar las características que no son relevantes en la clasificación de pacientes y mejorar los resultados de las predicciones de los algoritmos. Para esto, se tomaron los resultados obtenidos de los análisis de correlación de Pearson y Spearman, seleccionando únicamente las características que presentaron una correlación mayor que ± 0.6 con el diagnóstico clínico, esto se realizó para los dos grupos de datos por separado en los dos primeros casos de estudio, por lo que las características en cada caso no son las mismas. Para el tercer caso de estudio, se seleccionaron las características que presentaron una correlación mayor que ± 0.6 con la clasificación dada por el software PMOD. En los tres casos se eligió este valor umbral de correlación porque es el valor mínimo que se debe tener para evitar la predicción de una variable a partir de otra, sea aleatoria. Cabe mencionar que se utilizaron otros métodos de selección de características importantes como el Select K Best, Select from Model y el método Lasso, sin embargo los mejores resultados se obtuvieron con los dos análisis de correlación en todos los casos de estudio, por lo que solo se presentan los resultados con estos dos métodos. Se realizó nuevamente el entrenamiento de los diferentes modelos de aprendizaje autónomo supervisado, utilizando únicamente las características seleccionadas en cada caso y se realizaron las predicciones con los datos de prueba, utilizando de igual forma solo las características seleccionadas. Para todas las pruebas, los parámetros utilizados en los algoritmos

de aprendizaje autónomo se presentan en la tabla 2.1.

Tabla 2.1: parámetros importantes utilizados en cada algoritmo en el entrenamiento y pruebas.

Algoritmo	Parámetros
SVM	Kernel lineal
KNN	Distancia euclidiana, 5 vecinos
Red neuronal	Perceptrón multicapa (MLP), 100 capas ocultas
Regresión logística	Peso uniforme
Random Forest	100 árboles, 5 atributos por árbol

Se evaluó el rendimiento de cada algoritmo, en los tres casos de estudio, para las pruebas realizadas con todas las características, con las características seleccionadas mediante la correlación de Pearson y Spearman. Esto se realizó mediante el cálculo de la exactitud de clasificación (EC), la cual es simplemente una tasa entre las predicciones realizadas correctamente y el total de pacientes utilizados para las pruebas. También se calcularon la sensibilidad y especificidad, mediante las ecuaciones 2.4 y 2.5 [52].

$$\text{Sensibilidad} = \frac{VP}{VP + FN} \quad (2.4)$$

$$\text{Especificidad} = \frac{VN}{VN + FP} \quad (2.5)$$

En donde VP , FN , VN y FP son los verdaderos positivos, falsos negativos, verdaderos negativos y falsos positivos de las predicciones, respectivamente, tomando como positivos los clasificados como posible EA.

Adicionalmente, se realizó el entrenamiento de un algoritmo de clasificación por clústers denominado K-means, el cual permite clasificar un grupo de instancias, a partir de sus características, sin conocer previamente a qué clase pertenecen. Debido a esta característica, se considera este algoritmo como un modelo de aprendizaje no supervisado [34]. Para implementar este modelo se debe definir el número de clústers que se desean utilizar, es decir, en cuantas clases se desea clasificar al grupo de datos. El algoritmo funciona definiendo puntos denominados centroides, los cuales sirven para delimitar las regiones correspondientes a cada clúster. Para entender mejor este proceso, debemos pensar en un plano tridimensional en el que los ejes representan características de las instancias, cada instancia tiene una posición diferente en el espacio de características y el objetivo del algoritmo es definir a los centroides de manera tal que se formen grupos o clústers con instancias de características similares, como se muestra en la figura 2.8.

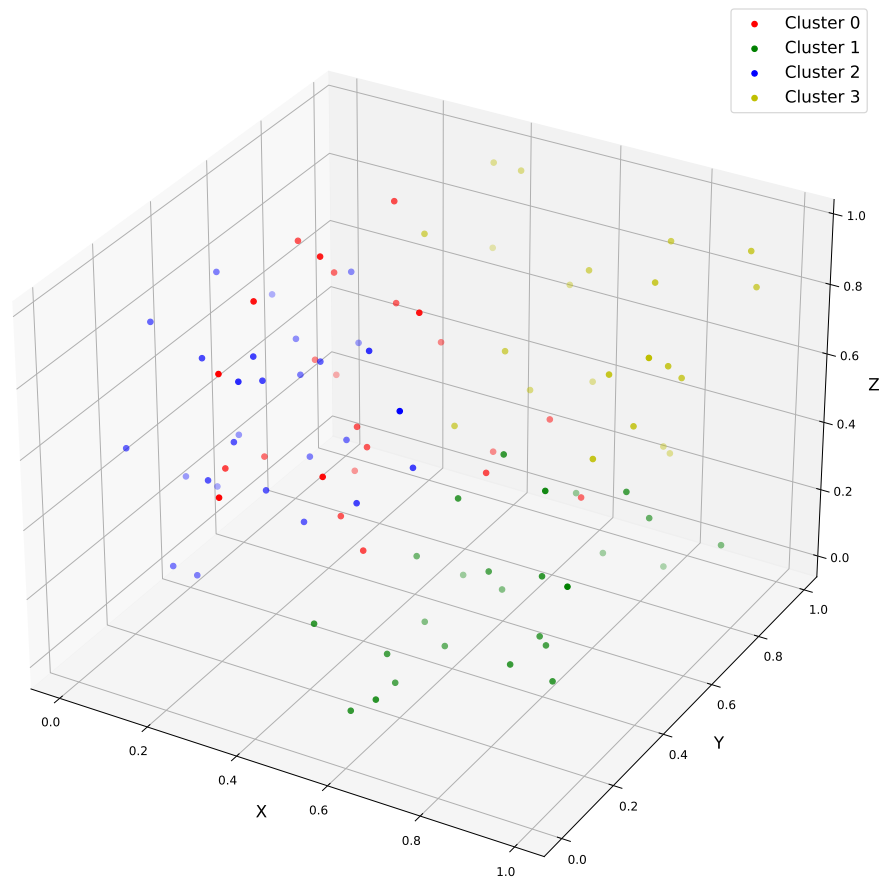


Figura 2.8: Visualización de la clasificación por clústers mediante el algoritmo K-means.

Cabe mencionar que en este ejemplo solo se consideran 3 características para poder visualizar los clústers, pero la clasificación se realiza considerando todas las características de las instancias. El entrenamiento y clasificación mediante este algoritmo, se realizó para los 3 casos de estudio, utilizando 2 clústers que fueron representativos de las dos clases de pacientes que se tienen en cada caso (Alzheimer y enfermedad neurológica ó Alzheimer y controles). Cabe mencionar que se utilizó el grupo completo de datos (50 pacientes) para el entrenamiento en cada caso de estudio, eliminando la clase objetivo y utilizando las mejores características obtenidas del análisis de correlación de Spearman. El rendimiento de este modelo se evaluó mediante el cálculo de la exactitud de su clasificación.

Capítulo 3

Resultados

A continuación se presentan los resultados más relevantes obtenidos en este trabajo.

3.1. Información demográfica y características de pacientes

Las características demográficas de los pacientes, cuyos estudios [^{18}F]FDG-PET fueron utilizados para este trabajo, incluyen su edad, género, peso y talla. También se cuenta con su diagnóstico clínico y con la clasificación obtenida mediante la herramienta para Alzheimer del Software PMOD, además del puntaje PET dado por esta misma. Las características de los 75 pacientes se presentan en la tabla 3.1.

Tabla 3.1: Información de los pacientes cuyos estudios PET fueron utilizados para este trabajo.

Dx	Puntaje PET	C. PMOD	Edad (años)	Género	Peso (kg)	Talla (m)
Alz	1.9123	1	80	0	64	1.56
Alz	3.0342	1	59	0	60	1.74
Alz	1.7641	1	77	1	54	1.54
Alz	1.8151	1	77	1	75	1.52
Alz	2.9227	1	73	1	69	1.5
Alz	1.6545	1	67	1	77	1.54
Alz	2.0705	1	75	1	75	1.53
Alz	1.6289	1	66	1	55	1.67
Alz	1.64	1	78	0	78	1.72
Alz	2.7882	1	70	1	43	1.53
Alz	1.7862	1	70	1	52	1.58
Alz	3.4961	1	69	1	61	1.52
Alz	1.9056	1	73	0	81	1.8

Tabla 3.1: Información de los pacientes cuyos estudios PET fueron utilizados para este trabajo.

Dx	Puntaje PET	C. PMOD	Edad (años)	Género	Peso (kg)	Talla (m)
Alz	0.9099	0	83	0	74	1.72
Alz	1.2819	1	78	0	52	1.59
Alz	0.4625	0	69	0	80	1.76
Alz	2.4114	1	83	1	57	1.58
Alz	0.9296	0	74	1	74	1.47
Alz	2.2506	1	73	1	40	1.52
Alz	1.1942	1	67	0	70	1.7
Alz	3.0921	1	75	0	68	1.78
Alz	3.7926	1	65	1	62	1.57
Alz	3.5587	1	57	0	69	1.65
Alz	2.6934	1	70	0	68	1.72
Alz	1.3128	1	65	1	68	1.58
PC	1.7149	1	73	1	60	1.5
EnfNeu	0.5038	0	92	1	59	1.59
EnfNeu	1.0678	1	67	0	69	1.71
EnfNeu	2.4141	1	74	1	47	1.67
EnfNeu	2.4566	1	66	0	61	1.58
EnfNeu	1.4441	1	86	1	50	1.55
EnfNeu	3.7376	1	70	0	66	1.65
EnfNeu	0.9291	0	66	0	102	1.78
EnfNeu	1.2039	1	77	1	76	1.56
EnfNeu	2.8467	1	66	0	70	1.72
EnfNeu	1.9405	1	54	1	90	1.6
EnfNeu	1.0072	1	73	1	34	1.44
EnfNeu	0.4487	0	55	1	34	1.5
EnfNeu	1.069	1	52	0	77	1.8
EnfNeu	1.5596	1	60	1	43	1.49
EnfNeu	0.944	0	70	0	83	1.72
EnfNeu	1.0951	1	56	1	54	1.52

Tabla 3.1: Información de los pacientes cuyos estudios PET fueron utilizados para este trabajo.

Dx	Puntaje PET	C. PMOD	Edad (años)	Género	Peso (kg)	Talla (m)
EnfNeu	0.6734	0	59	0	87	1.7
EnfNeu	0.5776	0	61	0	85	1.65
EnfNeu	1.8653	1	75	0	80	1.68
EnfNeu	2.9426	1	60	1	58	1.65
EnfNeu	1.0274	1	69	0	101	1.8
EnfNeu	1.2269	1	54	1	82	1.6
EnfNeu	1.4219	1	55	0	84	1.79
EnfNeu	0.9763	0	66	0	79	1.78
Con	0.03	0	66	1	56	1.54
Con	0.4379	0	60	1	61	1.48
Con	0.1498	0	77	0	55	1.7
Con	0.5246	0	59	1	69	1.5
Con	0.1716	0	61	1	58	1.5
Con	0.4631	0	53	1	60	1.54
Con	0.7407	0	65	1	52	1.44
Con	0.9895	0	61	0	139	1.85
Con	0.8083	0	88	1	69	1.59
Con	0.1353	0	65	0	72	1.67
Con	0.3854	0	72	1	46	1.45
Con	0.8016	0	68	0	77	1.71
Con	0.3869	0	69	0	76	1.62
Con	0.7601	0	58	0	57	1.72
Con	0.348	0	59	1	63	1.6
Con	0.4674	0	58	0	94	1.77
Con	0.2116	0	54	0	112	1.88
Con	0.1209	0	76	1	60	1.51
Con	0.3236	0	58	0	75	1.6
Con	0.4301	0	68	0	74	1.64
Con	0.4398	0	60	1	66	1.63

Tabla 3.1: Información de los pacientes cuyos estudios PET fueron utilizados para este trabajo.

Dx	Puntaje PET	C. PMOD	Edad (años)	Género	Peso (kg)	Talla (m)
Con	0.4371	0	54	0	83	1.76
Con	0.0904	0	56	1	57	1.51
Con	0.9139	0	71	0	101	1.7
Con	0.1272	0	53	1	64	1.61

Dx: diagnóstico clínico, C. PMOD: clasificación de PMOD, Alz: Alzheimer, EnfNeu: Enfermedad Neurológica, Con: Controles, Género(1): Femenino, Género(0): Masculino.

Se puede notar que de los 25 pacientes con EA en su diagnóstico clínico, 23 fueron clasificados como probable Alzheimer por la herramienta del software PMOD, por lo que debemos considerar en el tercer caso de estudio que se tienen 23 pacientes con probable EA y 27 controles, ya que la clase objetivo está dada por la clasificación del software PMOD en este caso. Cabe mencionar que los pacientes que tienen indicado Alzheimer en su diagnóstico clínico son en algunos casos pacientes con un estado avanzado de la enfermedad y en otros son pacientes que llegan con una alta sospecha de padecimiento de la EA fundamentada por pruebas neuropsicológicas, y que además presentan el patrón metabólico característico de la enfermedad según los médicos radiólogos de la unidad. Los pacientes que tienen indicado Alzheimer en su diagnóstico clínico, también pueden ser referidos como sospechosos de EA en este trabajo. Los pacientes con alguna enfermedad neurológica cuentan con estudios previos que sustentan los diagnósticos en cada caso. En la figura 3.1 se presenta la distribución de pacientes por edades que padecen alguna enfermedad neurológica, incluyendo EA, distinguiendo entre género masculino y femenino.

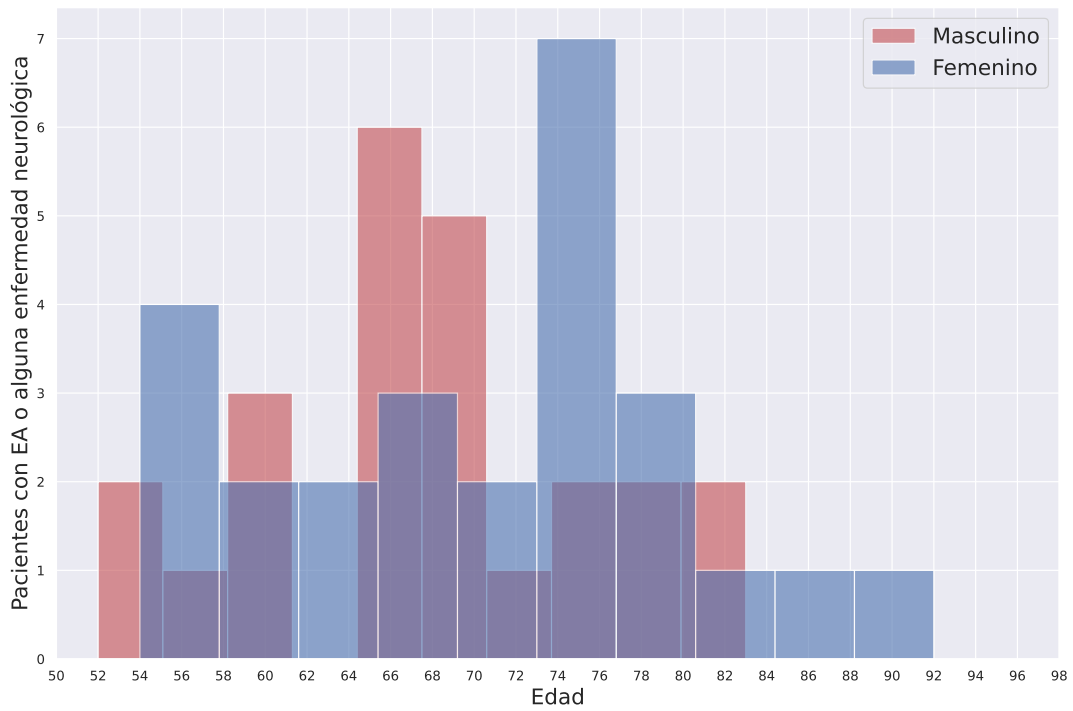


Figura 3.1: Histograma de la distribución de edades de pacientes con alguna enfermedad neurológica o con EA, distinguiendo entre género masculino y femenino.

También es ilustrativo comparar las distribuciones de edades de los diferentes tipos de pacientes de acuerdo con su diagnóstico clínico, como se presenta en la figura 3.2.

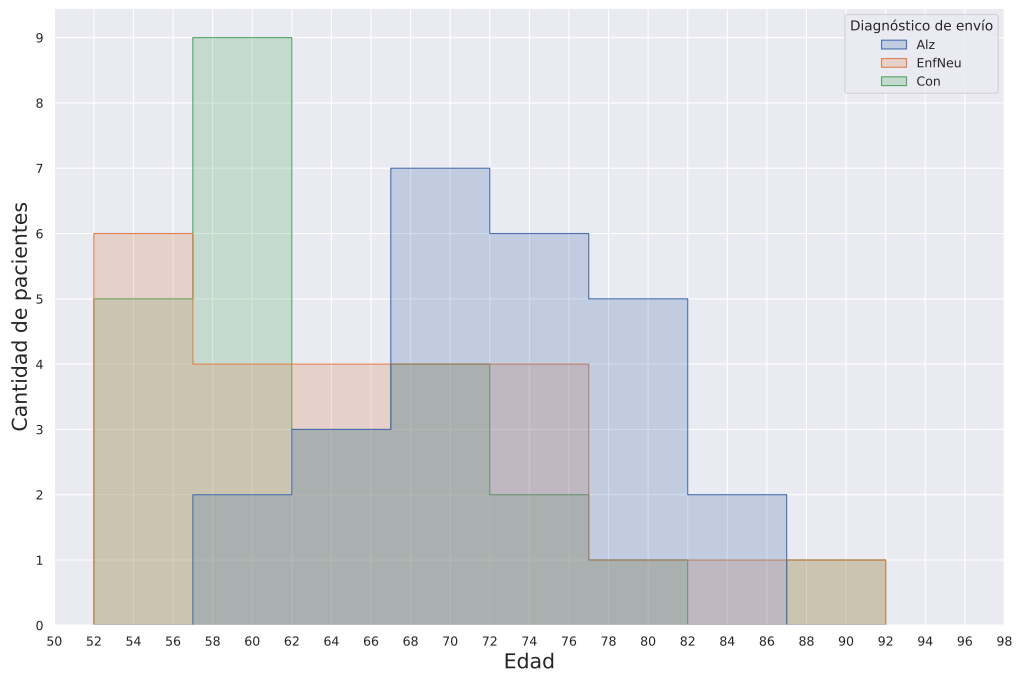


Figura 3.2: Histograma de la distribución de edades distinguiendo entre pacientes con EA (Alz), con enfermedad neurológica (EnfNeu) y controles (Con), de acuerdo con su diagnóstico clínico.

3.2. Imágenes [^{18}F]FDG-PET

A continuación se presenta como ejemplo uno de los estudios [^{18}F]FDG-PET. Únicamente se muestran una serie de cortes axiales del estudio de uno de los pacientes con EA indicada en su diagnóstico clínico.

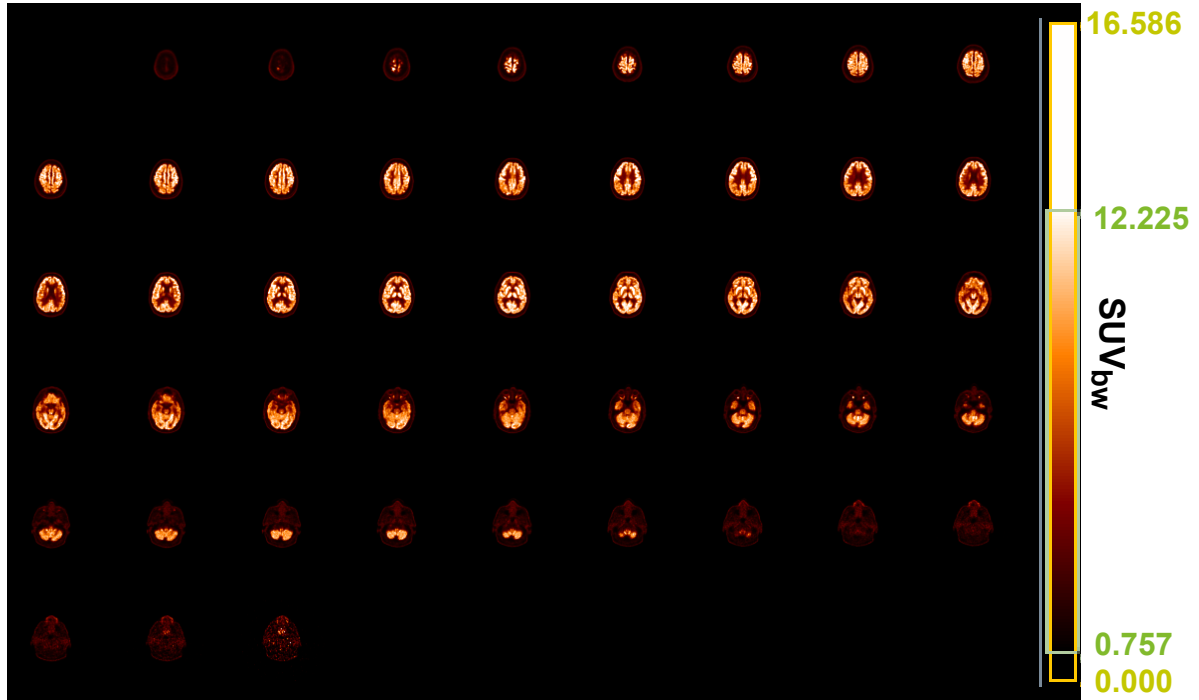


Figura 3.3: Serie de cortes axiales de estudio [^{18}F]FDG-PET cerebral de paciente con EA en su diagnóstico clínico, en escala SUV_{bw} .

La imagen está en escala SUV_{bw} , que es lo mismo que el SUV pero mediante el subíndice bw se especifica que se calculó con base en el peso del paciente.

3.3. Mapas de t valores

En la figura 3.4 se muestran dos grupos de imágenes que contienen cortes axiales, sagitales y coronales de dos pacientes, uno con EA en su diagnóstico clínico y el otro es un paciente control. Los cortes en ambos estudios son los mismos, las imágenes son estudios estándar de RM de secuencia T1 fusionadas con mapas de t valores, los cuales se obtuvieron luego de aplicar la máscara con las regiones cerebrales correlacionadas con la EA de acuerdo con el MMSE a las imágenes en escala de t valores de ambos pacientes.

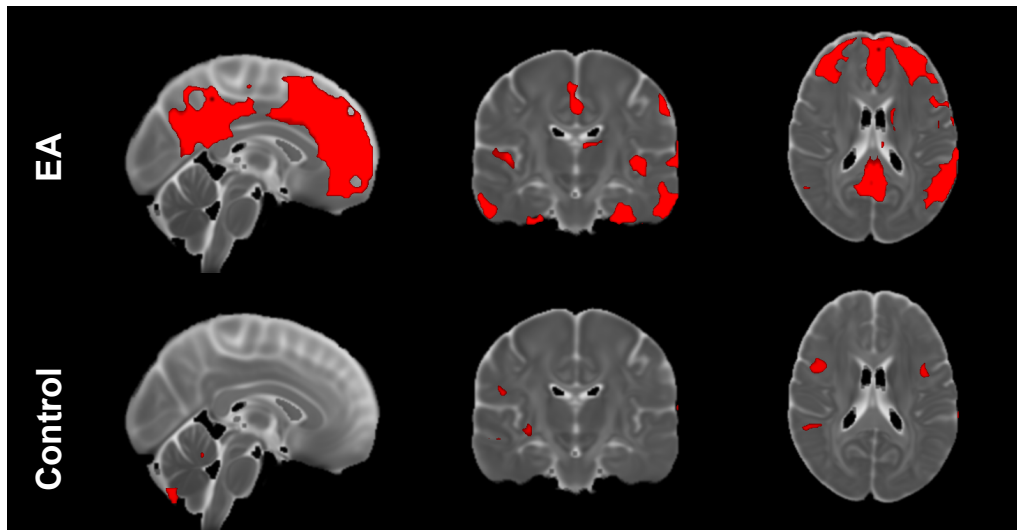


Figura 3.4: Imágenes estándar de resonancia magnética, secuencia T1, fusionadas con mapa de t valores de paciente con EA en su diagnóstico clínico y paciente control. La escala de ambas imágenes es arbitraria. Imágenes fusionadas mediante el software FSL [53].

Los niveles de intensidad de las imágenes se muestran de forma arbitraria, ya que únicamente se visualizan las regiones correlacionadas con la EA obtenidas luego de aplicar la máscara con los voxels correlacionados con la EA de acuerdo con el MMSE, dichas regiones se encuentran saturadas. Estas imágenes nos permiten observar la diferencia en la cantidad de regiones implicadas en la EA que pueden llegar a tener t valores positivos en pacientes con cada diagnóstico. Para observar de mejor manera la diferencia en la cantidad de regiones con alteración en ambos pacientes y su ubicación en el cerebro, en la figura 3.5 se presenta la fusión de los mapas de t valores con la imagen estándar, secuencia T1, de RM en diferentes vistas tridimensionales.

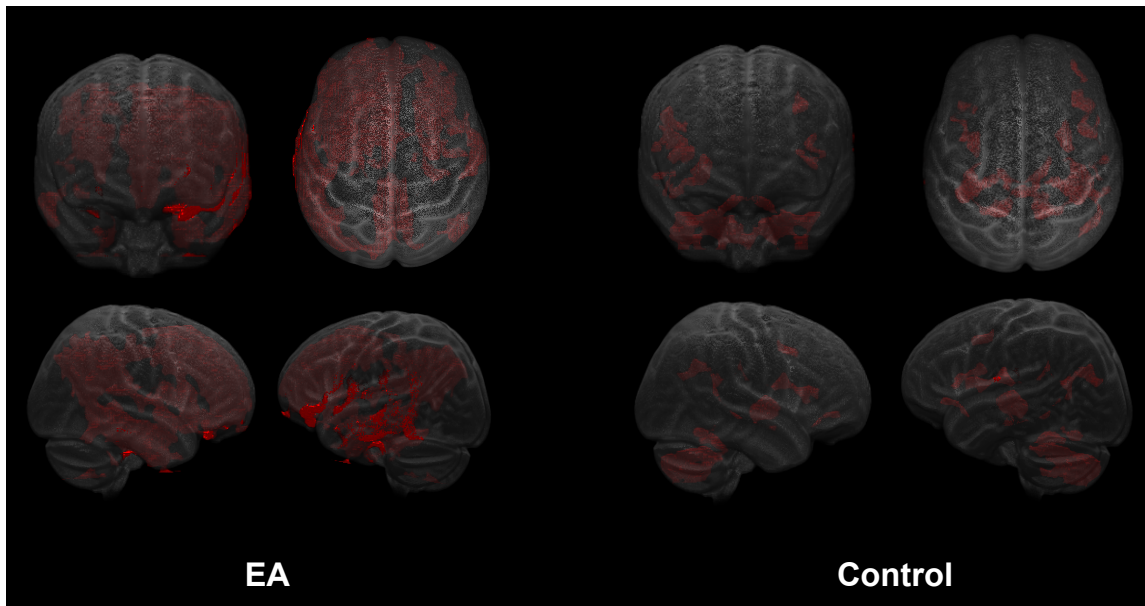


Figura 3.5: Mapa t 3D fusionado con imagen estándar de resonancia magnética con secuencia T1 de paciente con EA en su diagnóstico clínico y paciente control. La escala es arbitraria. Imágenes fusionadas mediante el software FSL [53].

De nueva cuenta se tiene una escala arbitraria, se disminuyó la opacidad de la imagen de RM para poder apreciar las regiones anormales con respecto a la EA.

3.4. Análisis de correlación

Se calculó el coeficiente de correlación de Pearson entre cada una de las características de radiómica medidas en las diferentes regiones cerebrales dadas por el atlas, con la finalidad de observar si existe relación entre las mismas características y para seleccionar las que mostraban mayor relación con la clasificación de cada paciente. Se seleccionaron las características que presentaron un coeficiente de correlación con el diagnóstico clínico mayor que ± 0.6 , con la finalidad de utilizarlas para realizar una de las pruebas con los algoritmos de aprendizaje autónomo. Esto se realizó para los dos primeros casos de estudio descritos en la figura 2.5 y para el tercer caso de estudio se seleccionaron las características que presentaron un coeficiente de correlación mayor que ± 0.6 con la clasificación dada por PMOD. También se calculó el coeficiente de correlación de Spearman para todas las características y de igual forma se seleccionaron las que presentaron coeficientes mayores que ± 0.6 con el diagnóstico clínico o con la clasificación de PMOD, según sea el caso de estudio. En la figura 3.6 se observa la matriz de correlaciones de Spearman con las características seleccionadas en el primer caso de estudio.

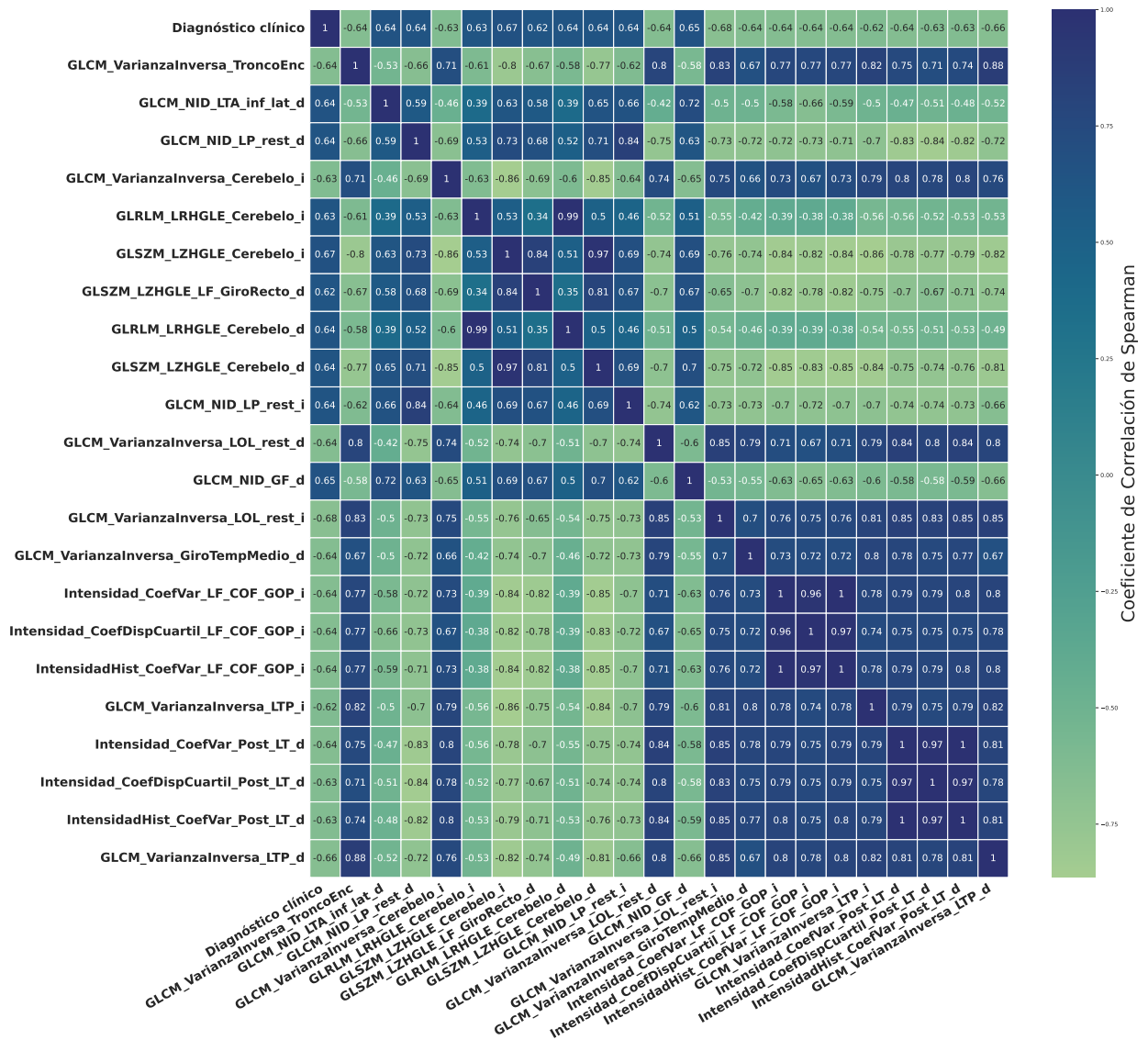


Figura 3.6: Mapa de calor de la matriz de correlaciones de Spearman con las características más correlacionadas con el diagnóstico clínico para el primer caso de estudio.

En la figura 3.7 se presentan las características con una mayor correlación con el diagnóstico clínico para el segundo caso de estudio, es decir, para el grupo conformado por pacientes con EA y controles.

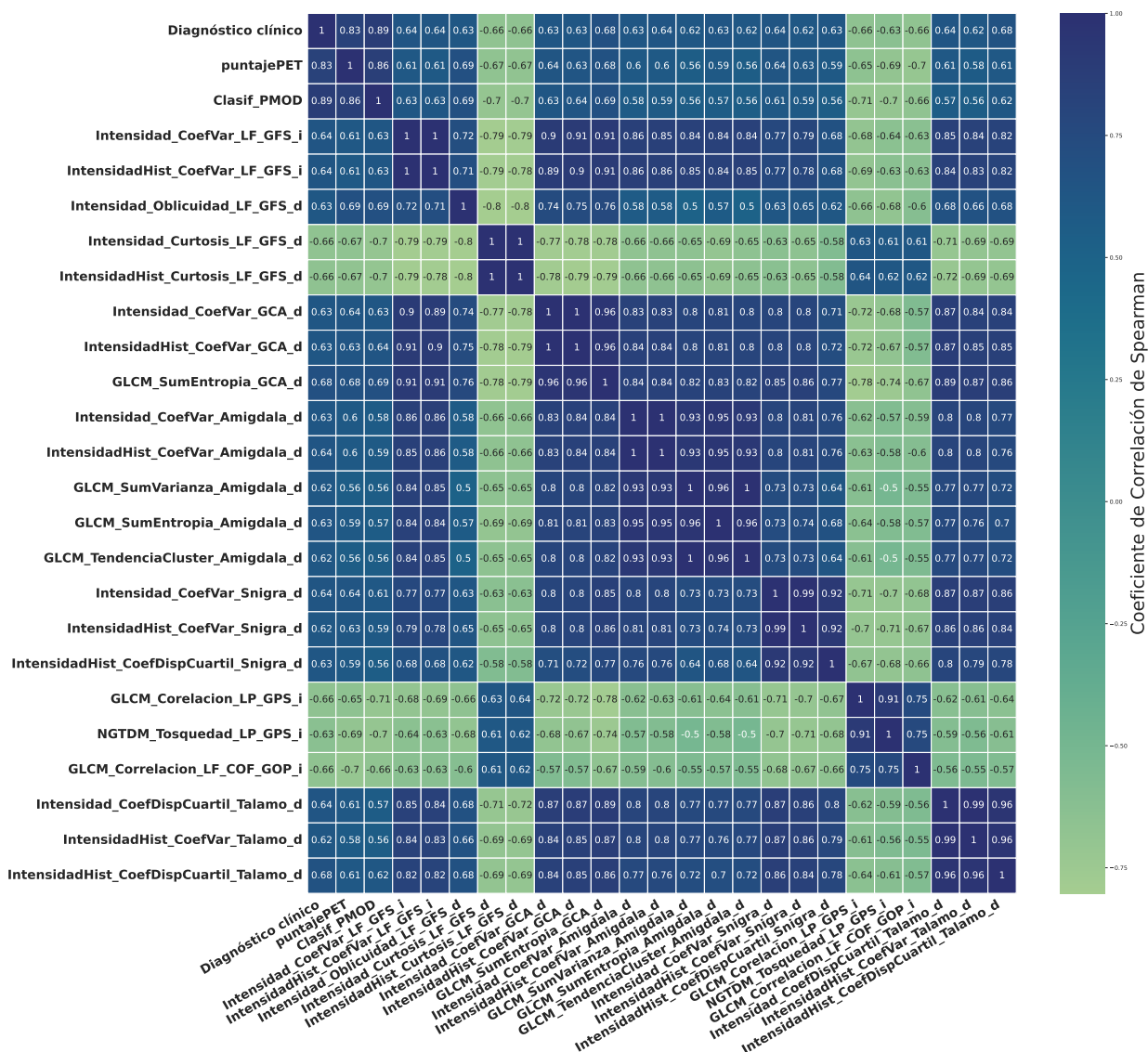


Figura 3.7: Mapa de calor de la matriz de correlaciones de Spearman con las características más correlacionadas con el diagnóstico clínico para el segundo caso de estudio.

Se puede notar que las características que presentan una mayor correlación para ambos casos de estudio son diferentes. Esto mismo ocurre en el tercer caso de estudio, donde se eligieron las características que presentaron un coeficiente de correlación de Spearman con la clasificación de PMOD mayor, como se observa en la figura 3.8.

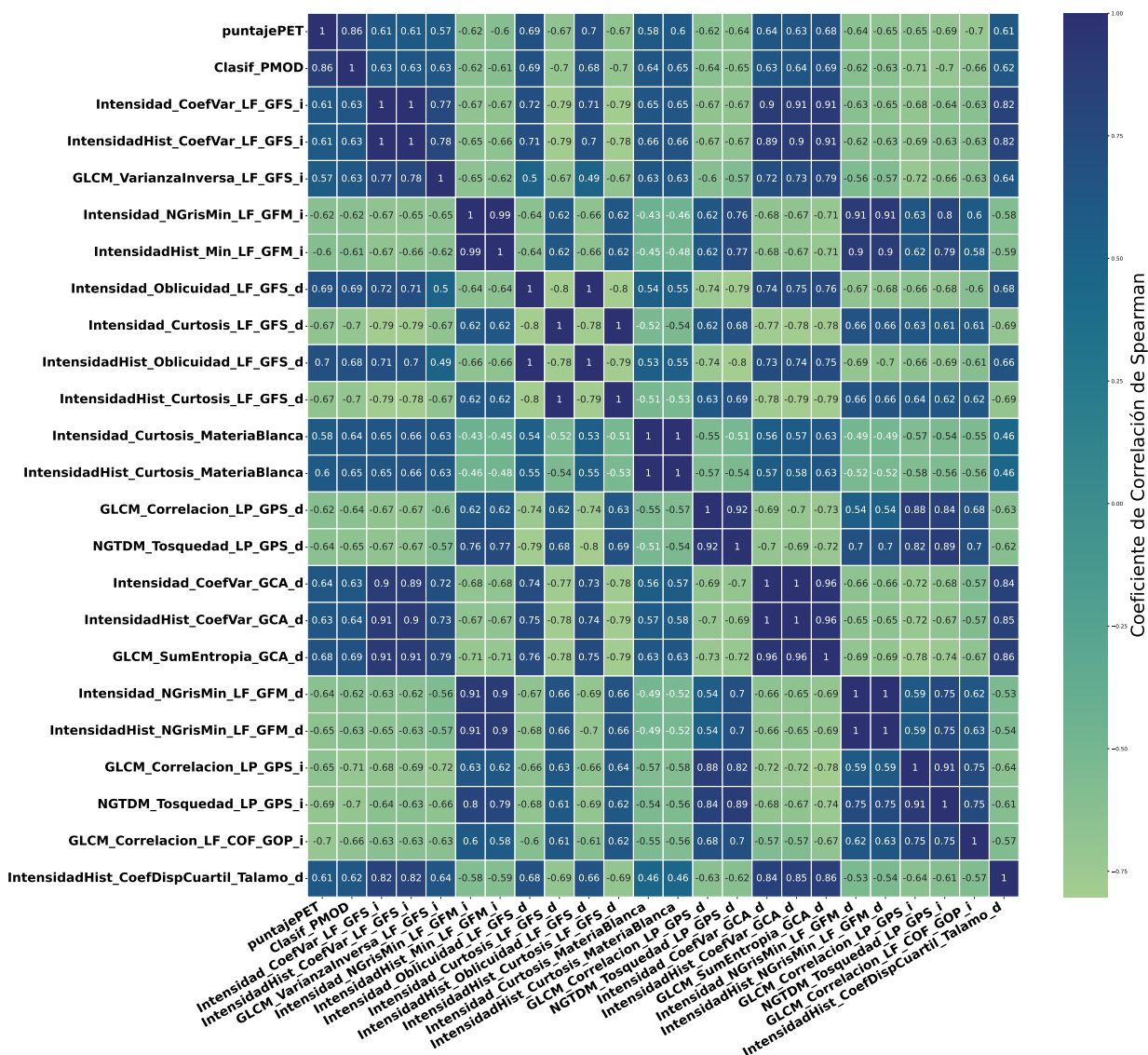


Figura 3.8: Mapa de calor de la matriz de correlaciones de Spearman con las características más correlacionadas con la clasificación dada por PMOD (Clasif PMOD) para el tercer caso de estudio.

Se puede notar que a pesar de tener a los mismos pacientes del segundo caso de estudio, las características obtenidas son diferentes debido a que la clasificación de PMOD no es exactamente igual que la dada por el diagnóstico clínico y esto puede deberse a que el software no es completamente preciso. En la figura 3.9 se muestran diagramas de caja de 3 características por cada caso de estudio que presentaron los mayores coeficientes de correlación de Spearman.

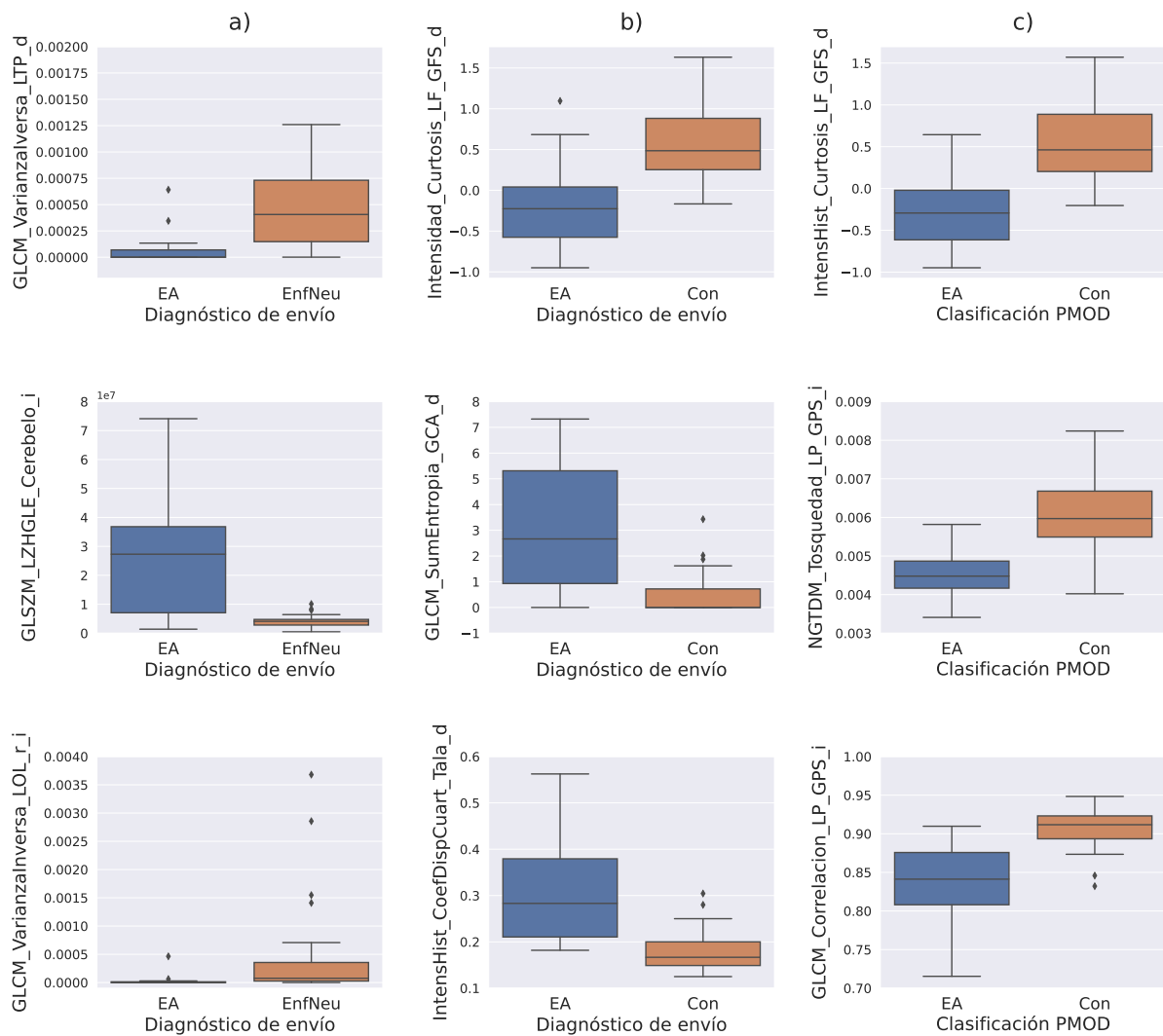


Figura 3.9: Diagramas de caja de las características con mayores coeficientes de correlación de Spearman del a) primer caso (Alz - EnfNeu), b) segundo caso (Alz-Con) y c) tercer caso (Alz-Con PMOD) de estudio.

Se puede notar que en todos los casos es posible diferenciar visualmente los valores de las características de acuerdo con el tipo de paciente que se tiene, sin embargo, es necesario realizar una prueba cuantitativa para poder afirmar que ambos tipos de pacientes tienen valores diferentes. También se observa que en el primer caso de estudio, las características seleccionadas parecen tener un mismo valor para un tipo de paciente, lo cual se discutirá más adelante. Las definiciones de las características radiómicas se pueden encontrar en los anexos y las abreviaturas de las regiones cerebrales se pueden encontrar en el índice de abreviaturas.

3.5. Prueba de Mann-Whitney U

Se realizó la prueba de Mann-Whitney U, comparando las características que obtuvieron mayores coeficientes de correlación de Spearman en cada caso de estudio, para comprobar que los valores de estas son diferentes en cada tipo de paciente. En la tabla 3.2 se muestran los

valores p y del estadístico de prueba para cada una de las características mostradas en la figura 3.9.

Tabla 3.2: Resultados de la prueba de Mann-Whitney U para las características con mayor coeficiente de correlación de Spearman en cada caso de estudio.

Primer caso de estudio (Alz-EnfNeu)			
Característica	Zona	Valor p	Estadístico
GLCM_VarianzaInversa	Lóbulo temporal anterior derecho	4.40×10^{-6}	77
GLSZM_LZHGLE	Cerebelo Izquierdo	3.21×10^{-6}	553
GLCM_VarianzaInversa	Descanso del lóbulo occipital lateral izquierdo	2.26×10^{-6}	70.5
Segundo caso de estudio (Alz-Con)			
Característica	Zona	Valor p	Estadístico
Intensidad_Curtosis	Giro frontal superior derecho	3.53×10^{-6}	73
GLCM_SumEntropia	Giro del cíngulo anterior derecho	1.79×10^{-6}	557
IntensHist_CoefDispCuart	Tálamo derecho	2.35×10^{-6}	556
Tercer caso de estudio (Alz-Con PMOD)			
Característica	Zona	Valor p	Estadístico
IntensHist_Curtosis	Giro frontal superior derecho	1.19×10^{-6}	59
NGTDM_Tosquedad	Giro parietal superior izquierdo	8.86×10^{-7}	56
GLCM_Correlacion	Giro parietal superior izquierdo	8.02×10^{-7}	55

Se puede notar que todas las características cumplen con el criterio para afirmar que estas medidas son diferentes en ambos grupos de datos, en cada caso de estudio.

3.6. Predicciones utilizando todas las características

Luego de seleccionar los algoritmos que obtuvieron resultados aceptables (con un áreas bajo la curva ROC superiores a 0.7) en la prueba previa utilizando validación cruzada, se realizó el entrenamiento de dichos algoritmos para los tres casos de estudio descritos en la figura 2.5, utilizando todas las características de entrada que se presentaron en la figura 2.7. Luego del entrenamiento, se realizaron las pruebas (predicciones) con los algoritmos utilizando el 30% de los datos en cada caso (15 pacientes), en la tabla 3.3 se pueden observar los resultados de rendimiento en las predicciones de los 3 algoritmos que obtuvieron una EC mayor para cada caso.

Se observa que, en general, los resultados con mayor EC se obtienen con el algoritmo de redes neuronales. El área bajo la curva ROC es uno de los parámetros más utilizados para

Tabla 3.3: Resumen de los resultados de rendimiento obtenidos por los tres algoritmos con mayor exactitud en las predicciones en cada caso de estudio, utilizando todas las características de entrada.

Primer caso de estudio: pacientes con EA y enfermedad neurológica					
Algoritmos	AUC	EC	Sensibilidad	Especificidad	TE (s)
Red Neuronal	0.964	0.867	0.857	0.875	17.542
K Vecino más cercano	0.893	0.867	0.714	1.000	17.644
Máquinas de soporte vectorial	0.839	0.800	0.571	1.000	19.778
Segundo caso de estudio: pacientes con EA y controles					
Algoritmos	AUC	EC	Sensibilidad	Especificidad	TE (s)
Red Neuronal	0.907	0.800	0.800	0.756	16.850
K Vecino más cercano	0.556	0.667	0.667	0.556	18.696
Máquinas de soporte vectorial	0.500	0.733	0.733	0.600	18.038
Tercer caso de estudio: pacientes con EA y controles, clasif. de PMOD como objetivo.					
Algoritmos	AUC	EC	Sensibilidad	Especificidad	TE (s)
Red Neuronal	0.833	0.733	1.000	0.556	14.675
Random Forest	0.907	0.733	0.833	0.667	4.437
Máquinas de soporte vectorial	0.722	0.667	0.833	0.556	18.656

EC: exactitud de clasificación. AUC: area bajo la curva. TE: Tiempo de entrenamiento.

determinar si un modelo de clasificación es bueno, el cálculo de este valor se puede entender de mejor manera gráficamente en la figura 3.10.

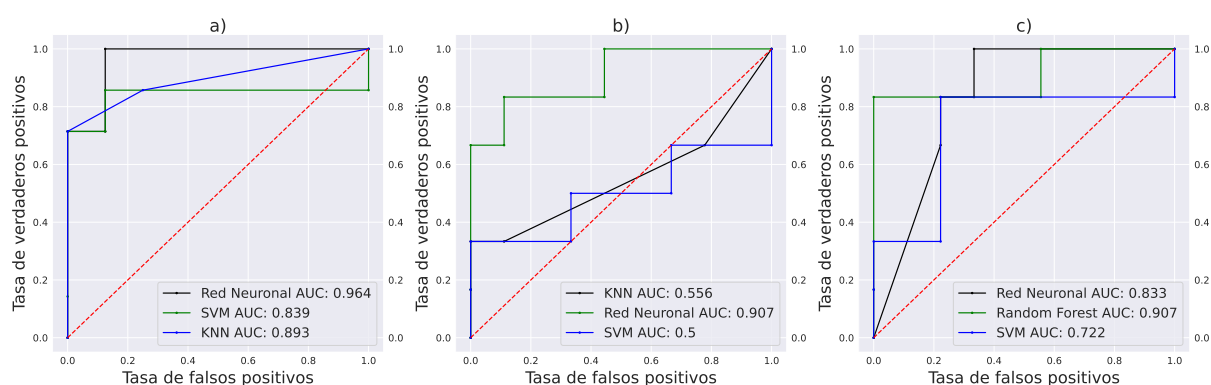


Figura 3.10: Curvas ROC del a) primer caso (Alz - EnfNeu), b) segundo caso (Alz-Con) y c) tercer caso (Alz-Con PMOD) de estudio, utilizando todas las características de entrada.

Se puede notar que las AUC calculadas corresponden a lo observado en las curvas ROC de cada clasificador. Cabe mencionar que la línea roja se denomina línea de no-discriminación y esta representa el límite mínimo para que las predicciones de un modelo sean aleatorias, es

decir, que se pueden considerar las predicciones de los modelos con una AUC inferior a esta línea como aleatorias.

3.7. Predicciones con las mejores características

Como se mencionó, mediante el cálculo del coeficiente de correlación de Pearson se seleccionaron las características con las que se obtuvieron coeficientes mayores que ± 0.6 con el diagnóstico clínico y clasificación de PMOD. Utilizando solo estas características para el entrenamiento de los algoritmos y pruebas en cada caso de estudio, se obtuvieron los resultados de rendimiento que se muestran en la tabla 3.4.

Tabla 3.4: Resumen de los resultados de rendimiento obtenidos por los tres algoritmos con mayor exactitud en sus predicciones en cada caso de estudio, utilizando características que tienen mayor correlación de Pearson con el diagnóstico clínico y clasificación de PMOD.

Primer caso de estudio: pacientes con EA y enfermedad neurológica					
Algoritmos	AUC	EC	Sensibilidad	Especificidad	TE (s)
Red Neuronal	0.944	0.867	1.000	0.778	0.135
Random Forest	0.944	0.867	1.000	0.778	0.273
Regresión Logística	0.944	0.867	1.000	0.778	0.021
Segundo caso de estudio: pacientes con EA y controles					
Algoritmos	AUC	EC	Sensibilidad	Especificidad	TE (s)
Red Neuronal	0.940	0.867	0.800	1.000	0.140
Regresión Logística	0.940	0.867	0.800	1.000	0.010
Random Forest	0.920	0.933	0.900	1.000	0.197
Tercer caso de estudio: pacientes con EA y controles, clasif. de PMOD como objetivo.					
Algoritmos	AUC	EC	Sensibilidad	Especificidad	TE (s)
Red Neuronal	0.946	0.733	0.857	0.625	0.162
Random Forest	0.911	0.800	0.857	0.750	0.262
Máquinas de soporte vectorial	0.946	0.800	0.857	0.750	0.075

EC: exactitud de clasificación. AUC: area bajo la curva. TE: Tiempo de entrenamiento.

De igual forma, solo se presentan los resultados de los algoritmos que obtuvieron una mayor exactitud de clasificación en cada caso de estudio. De manera general, se obtuvieron mejores resultados utilizando solo estas características y los tiempos de entrenamiento se redujeron considerablemente, lo cual es un factor que se debe tomar en cuenta si se desea utilizar esta metodología con una cantidad de pacientes mucho mayor. En la figura 3.11 podemos notar que hubo una mejoría en el AUC obtenida por los algoritmos con respecto a las pruebas utilizando todas las características.

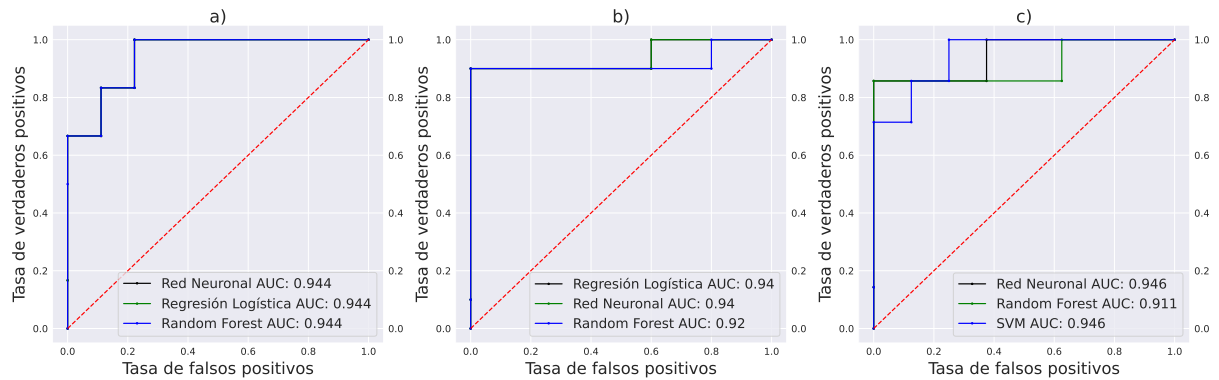


Figura 3.11: Curvas ROC del a) primer caso (Alz - EnfNeu), b) segundo caso (Alz-Con) y c) tercer caso (Alz-Con PMOD) de estudio, utilizando las características con un mayor coeficiente de correlación de Pearson.

Cabe mencionar que se utilizó el coeficiente de correlación de Pearson como referencia para la selección de características porque se obtuvieron mejores resultados en las predicciones para los tres casos de estudio, en contraste con los otros métodos de selección de características utilizados, pero el análisis por sí solo no debe ser considerado como válido para encontrar parámetros de referencia o marcadores de la EA a menos que se cuente con una métrica de carácter continuo relacionada con la severidad de la EA como el puntaje PET ya que, como se mencionó, el coeficiente de correlación de Pearson no es preciso cuando se tienen variables categóricas. De la misma manera, se seleccionaron las características con las que se obtuvo un mayor coeficiente de correlación de Spearman con el diagnóstico clínico o clasificación de PMOD y se utilizaron para realizar el entrenamiento y predicciones con los algoritmos en cada caso de estudio, los mejores resultados de rendimiento obtenidos se presentan en la tabla 3.5.

De nueva cuenta, podemos notar que hubo una mejoría en los resultados de las predicciones de los algoritmos, con respecto al uso del coeficiente de correlación de Pearson. Además, también se obtuvieron tiempos de entrenamiento bajos y AUC más altas con respecto a Pearson, como se observa en la figura 3.12.

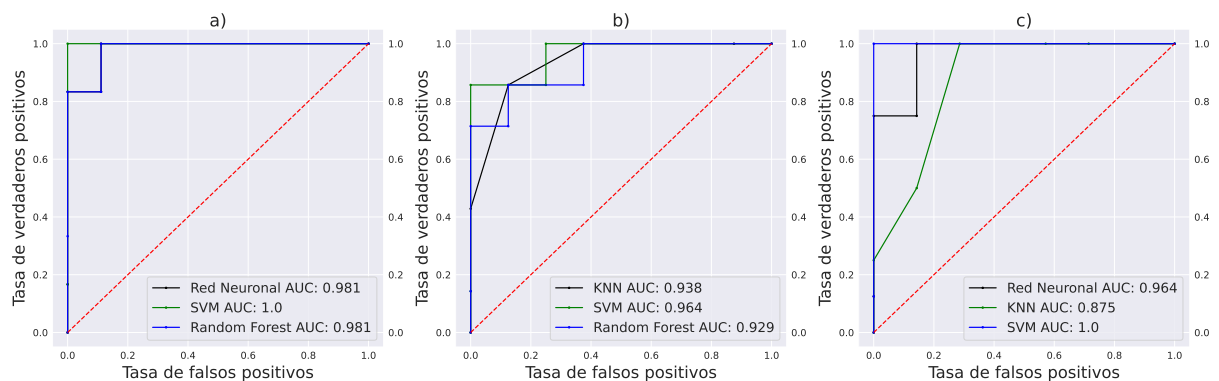


Figura 3.12: Curvas ROC del a) primer caso (Alz - EnfNeu), b) segundo caso (Alz-Con) y c) tercer caso (Alz-Con PMOD) de estudio, utilizando las características con un mayor coeficiente de correlación de Spearman.

Tabla 3.5: Resumen de los resultados de rendimiento obtenidos por los tres algoritmos con mayor exactitud en sus predicciones en cada caso de estudio, utilizando características que tienen mayor correlación de Spearman con el diagnóstico clínico y clasificación de PMOD.

Primer caso de estudio: pacientes con EA y enfermedad neurológica					
Algoritmos	AUC	EC	Sensibilidad	Especificidad	TE (s)
Red Neuronal	0.981	0.933	1.000	0.889	0.199
Random Forest	0.981	0.933	1.000	0.889	0.247
Máquinas de soporte vectorial	1.000	0.933	0.833	1.000	0.080
Segundo caso de estudio: pacientes con EA y controles					
Algoritmos	AUC	EC	Sensibilidad	Especificidad	TE (s)
K Vecino más cercano	0.938	0.867	0.857	0.875	0.081
Máquinas de soporte vectorial	0.964	0.867	0.857	0.875	0.083
Random Forest	0.929	0.800	0.857	0.750	0.241
Tercer caso de estudio: pacientes con EA y controles, clasif. de PMOD como objetivo.					
Algoritmos	AUC	EC	Sensibilidad	Especificidad	TE (s)
Red Neuronal	0.964	0.933	1.000	0.857	0.165
Máquinas de soporte vectorial	1.000	0.933	1.000	0.857	0.074
K Vecino más cercano	0.875	0.867	1.000	0.714	0.074

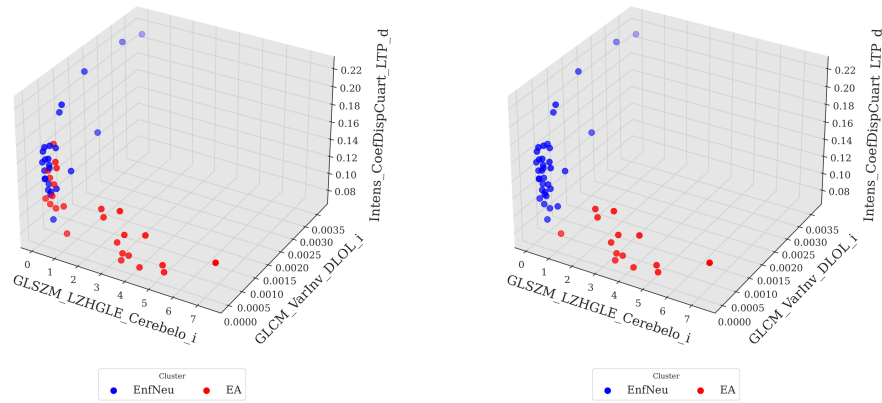
EC: exactitud de clasificación. AUC: area bajo la curva. TE: Tiempo de entrenamiento.

Podemos notar de igual forma que con el algoritmo de redes neuronales se siguen obteniendo buenos resultados, pero el algoritmo que obtiene el mejor rendimiento en las predicciones de todos los casos de estudio utilizando las características con mayor coeficiente de correlación de Spearman es el de máquinas de soporte vectorial, lo cual se discutirá en el siguiente capítulo.

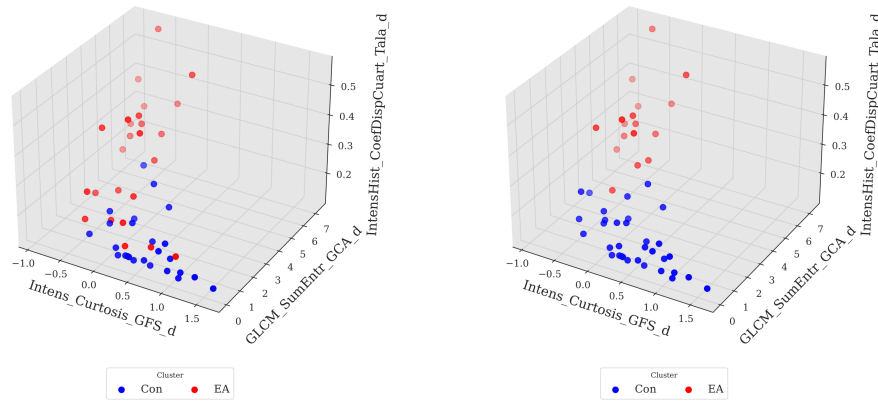
3.8. Clasificación por clústers

Como se mencionó, en este caso se utilizó el grupo completo de datos para cada caso de estudio y se utilizaron las características obtenidas en el análisis de correlación de Spearman. Se seleccionaron dos clústers para clasificar a los pacientes y se comparó la clasificación dada por el algoritmo kmeans con la clase objetivo en cada caso de estudio. En la figura 3.13 se observa de manera gráfica la clasificación realizada por el algoritmo (derecha) y la clasificación verdadera de los pacientes (izquierda) dada por la clase objetivo, en cada caso de estudio. En las gráficas solo se muestran algunas características utilizadas en los ejes para poder visualizar la clasificación, sin embargo, el algoritmo utiliza todas las características para formar los clústers, como se mencionó anteriormente. Para el primer caso de estudio, se obtuvo una EC, utilizando K-means, del 78 %, para el segundo caso de estudio se obtuvo una EC del 80 % y para el tercer caso se obtuvo una EC del 86 %.

a)



b)



c)

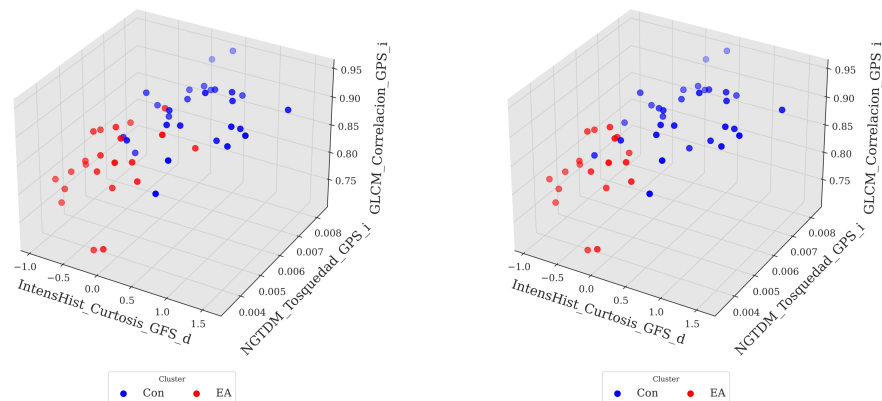


Figura 3.13: Representación gráfica de la clasificación por clústers de K-means (derecha) y la clasificación verdadera (izquierda) dada por la clase objetivo en el a) primer caso (Alz - EnfNeu), b) segundo caso (Alz-Con) y c) tercer caso (Alz-Con PMOD) de estudio, utilizando las características con un mayor coeficiente de correlación de Spearman.

Capítulo 4

Discusión

En este capítulo se discuten los aspectos más importantes de los resultados obtenidos. Las enfermedades neurodegenerativas suelen presentarse desde los 60 años, aumentando hasta los 70 y presentando una disminución de los casos en edades más avanzadas debido a la esperanza de vida. En esta muestra no parece haber una diferencia significativa de los casos en cuanto al género, sin embargo, la muestra es pequeña para realizar cualquier afirmación, algunos estudios han mostrado una mayor incidencia de este tipo de padecimientos en el género femenino debido a cuestiones genéticas y hormonales [54]. En cuanto a la distribución por edades, distinguiendo entre los 3 tipos de pacientes que se tienen en el estudio (pacientes con sospecha de EA, pacientes con alguna enfermedad neurológica y controles), se puede notar en la figura 3.2 que los pacientes con sospecha de EA siguen la misma distribución esperada, mientras que los pacientes con alguna enfermedad neurológica disminuyen gradualmente desde los 50 años. De nueva cuenta se debe mencionar que la muestra de pacientes es pequeña debido a la limitada cantidad de estudios PET neurológicos que cumplen con los criterios de inclusión de este estudio, no obstante, se espera que la cantidad de este tipo de estudios continúe en aumento.

En cuanto a los mapas de t valores obtenidos mediante la herramienta de Alzheimer del software usado, se observa una clara diferencia entre una imagen de paciente con sospecha de EA y un control en la figura 3.4 ya que, las partes rojas en las imágenes, las cuales representan las regiones cerebrales correlacionadas con el MMSE que tuvieron t valores positivos, son más abundantes en el paciente con sospecha de EA. Esto conlleva a que la suma de los t valores y por lo tanto el puntaje PET sean mayores. En la figura 3.5 también se aprecia esta clara diferencia entre ambos pacientes y además en el paciente con sospecha de EA se observa lateralidad en la cantidad de regiones anormales con respecto a la EA. Concretamente, se obtuvieron más de estas regiones en el lado izquierdo del cerebro del paciente. La cantidad de regiones que permanecen en la imagen luego de aplicar la máscara que contiene las regiones correlacionadas con el MMSE y por lo tanto con la EA [43], está dada por la cantidad de estas regiones que tengan una actividad o captación menor que la esperada de acuerdo con su edad. Debido a esto, entre más regiones con menor captación se tengan, mayor será el puntaje PET, recordemos que este valor es el que utiliza el software para clasificar a los pacientes. Si bien esta puede ser una buena estimación, la herramienta tiene deficiencias y por lo tanto también puede tener errores, por ejemplo, si se tiene un paciente con sospecha de EA pero además tiene alguna otra afectación no relacionada con la enfermedad en alguna de las regiones de la máscara, que no ha

sido identificada con anterioridad, el programa lo podría clasificar como posible EA debido a un padecimiento diferente. Por este motivo, el programa tiene la desventaja de solo funcionar adecuadamente con pacientes con síntomas de EA y controles. Además, el programa reporta que su herramienta tiene sensibilidades/especificidades de 83%/78% y 78%/94% con diversas bases de datos [55]. A pesar de estos factores, la clasificación dada por el programa fue utilizada como referencia en el tercer caso de estudio debido a que en este grupo de pacientes solo se utilizaron controles y pacientes con sospecha de EA.

Los resultados del análisis de correlación de Spearman en el primer caso de estudio, muestran que la mayoría de características que presentan una correlación superior a ± 0.6 con el diagnóstico clínico fueron medidas en regiones cerebrales que normalmente presentan hipometabolismo en la EA, como se observa en la figura 3.6. No obstante, algunas de estas características fueron medidas en regiones que no forman parte del patrón hipometabólico típico de la enfermedad, tales como el tronco encefálico y el cerebelo. En el caso del tronco encefálico, se han encontrado cambios morfológicos en esta región en estudios post mortem de pacientes con EA y se sugiere que esto ocasiona algunos de los síntomas de la EA [56], incluso se ha propuesto como biomarcador en conjunto con estudios de imagen [57]. El cerebelo por su parte, históricamente no se considera como una zona afectada dentro de la fisiopatología de la EA, sin embargo, recientemente se han realizado estudios en modelos animales en los cuales se ha encontrado presencia de las placas seniles en el cerebelo incluso en etapas tempranas de la enfermedad [58], por lo que se ha planteado que también sea una zona afectada en la EA ya que cumple funciones de aprendizaje, además de motoras [59]. Las características con un coeficiente de correlación de Spearman mayor en el primer caso de estudio, las cuales se presentan en la figura 3.9, fueron medidas en el lóbulo temporal posterior (LTP) derecho, en el descanso del lóbulo occipital lateral (LOL) derecho y en el cerebelo derecho. La característica medida en el descanso del LOL y en el LTA es una medida de homogeneidad, es decir que mientras mayor sea su valor, mayor es la homogeneidad en la región. En los diagramas de caja correspondientes a estas características se puede observar que, en estas dos regiones, los pacientes con sospecha de EA parecen tener poca o nula homogeneidad en la captación de $[^{18}\text{F}]\text{FDG}$ en estos tejidos, lo cual podría estar relacionado con la atrofia que sufren estas regiones en pacientes con EA. Los pacientes con alguna enfermedad neurológica, en general, parecen tener mayor homogeneidad en la captación del radiofármaco. En el cerebelo, la característica $GLSZM_{LZHGLE}$ representa el énfasis en los niveles de gris altos en regiones largas, es decir que mientras mayor sea este valor, mayor será la cantidad de niveles de gris iguales en zonas grandes dentro de la región. En el diagrama de caja correspondiente a esta característica, se puede notar que los pacientes con sospecha de EA presentan una mayor cantidad de zonas grandes con niveles de gris altos iguales en el cerebelo derecho, en comparación con los pacientes con alguna enfermedad neurológica. Esto sugiere que en el cerebelo de pacientes con sospecha de EA podrían haber zonas de mayor captación homogéneas, en comparación con los pacientes neurológicos, lo cual podría indicar que esta región no está siendo afectada en la EA a diferencia de en las otras enfermedades, aunque también podríamos pensar que en el cerebelo de pacientes con EA se tienen muchas zonas afectadas y otras zonas permanecen con la misma captación lo cual aumenta el valor de esta característica al tener más zonas con niveles de gris altos iguales. Al observar los diagramas de caja en este caso de estudio, se puede notar que los grupos de pacientes parecen tener una diferencia importante ya que los valores de estas características parecen tener distribuciones muy diferentes. Además de las posibles causas fisiológicas y anatómicas ya mencionadas, otra posible explicación sería

que los estudios de pacientes con alguna enfermedad neurológica, en general, son estudios más recientes, esto podría sugerir que la diferencia se debe a que algunos estudios se realizaron en equipos diferentes ya que la unidad PET/CT de la UNAM adquirió un equipo nuevo en el año 2019. Sin embargo, el grupo de pacientes con sospecha de EA también cuenta con 10 estudios recientes, 10 estudios realizados entre 2017 y 2022, y 5 estudios realizados antes de 2017, por lo que la diferencia entre los estudios realizados en ambos equipos no debería ser significativa para que se presenten las diferencias en las distribuciones de valores de estas características entre los dos tipos de pacientes que se muestran en los diagramas de caja. Además, las diferencias en los valores de intensidad debidas al uso de equipos distintos se reducen al hacer el cálculo de las características radiómicas, ya que se realiza un proceso de discretización de niveles de gris a un valor fijo de 64 niveles.

En el segundo caso de estudio también se encontraron algunas medidas correlacionadas con el diagnóstico clínico, como se observa en la figura 3.7, pertenecientes a regiones que no corresponden al patrón hipometabólico típico de la enfermedad: el tálamo derecho, la sustancia nigra derecha y la amígdala derecha. El tálamo es una estructura cerebral que cumple funciones de procesamiento de información sensorial, motora y de la memoria. Esta estructura tampoco ha sido relacionada con la EA, sin embargo, algunos estudios sugieren que esta estructura también podría verse afectada en etapas tempranas de la enfermedad por las placas seniles [60]. Lo mismo ocurre con la amígdala, la cual cumple diversas funciones entre las que se encuentra la modulación de la memoria episódica. Se ha encontrado evidencia de la presencia de placas seniles en estudios post mortem de pacientes con EA en la amígdala y se han realizado estudios que sugieren que esta región sufre de una atrofia comparable a la del hipocampo en etapas tempranas de la enfermedad [61]. En el caso de la sustancia nigra, esta es una estructura ubicada en el mesencéfalo que cumple funciones de aprendizaje y normalmente se asocia con la fisiopatología de la enfermedad de Parkinson, pero algunos estudios han mostrado que la sustancia nigra también se ve afectada por los NFT's en pacientes con EA [62]. Las características con mayor correlación de Spearman fueron medidas en regiones del giro frontal superior (GFS) derecho, giro del cíngulo anterior (GCA) derecho y en el tálamo derecho. El GFS se encuentra dentro de las regiones del patrón hipometabólico típico de la EA y, como se explicó anteriormente, el tálamo también podría estar implicado en las regiones afectadas en etapas tempranas de la enfermedad. El giro del cíngulo posterior es una de las regiones que forman parte del patrón hipometabólico de la EA, sin embargo, el GCA típicamente no se encuentra dentro de este patrón. No obstante, diversos estudios han demostrado que el GCA se ve afectado durante la EA y también se ha encontrado una reducción en el tamaño de esta región como un precedente para la evolución del deterioro cognitivo leve a la EA [63]. A pesar de esto, no se suele considerar dentro del patrón típico ya que esta región también se ve afectada en la demencia frontotemporal (DFT) [21] y, por lo tanto, un daño en la región del GCA es poco específico. En el primer caso de estudio, no se obtuvo alguna correlación significativa entre alguna característica medida en la región del GCA y el diagnóstico clínico, lo cual puede ocurrir debido a que dentro de los pacientes con enfermedad neurológica hay casos de DFT. La característica que se midió en el GFS derecho fue la de curtosis, la cual como se mencionó es un indicativo de la forma de la distribución de niveles de gris en la región, específicamente de qué tan angosta o puntiaguda es. En el diagrama de caja correspondiente a la característica, se observa que los pacientes con sospecha de EA tienen una distribución menos angosta, lo cual nos puede indicar que se tienen muchas zonas con diferente captación del radiofármaco, ó que las zonas que normalmente tienen una

captación promedio del mismo se redujeron, lo cual es consistente con el patrón hipometabólico de la EA. La característica medida en el GCA derecho con una mayor correlación fue la $GLCM_{SumEntropia}$, la cual es una medida de la aleatoriedad dentro de la matriz de co-ocurrencia de niveles de gris (definida en los anexos de este trabajo), esto significa que mientras mayor sea su valor, mayor será la aleatoriedad de zonas con niveles de gris similares. En su diagrama de caja correspondiente, se puede notar que los pacientes con sospecha de EA, en general, presentan valores mayores de este parámetro comparados con los controles en la región del GCA, esto puede deberse a la aleatoriedad de las zonas con atrofas que las placas seniles podrían estar ocasionando en dicha región. En cuanto al tálamo derecho, la característica medida con una mayor correlación de Spearman con el diagnóstico clínico fue la $IntensHist_{CoefDispCuart}$, la cual es una medida robusta de la dispersión del histograma de intensidades en la región. En su diagrama de caja correspondiente dentro de la figura 3.9, se observa que los pacientes con sospecha de EA presentan valores más altos de esta característica, es decir que tienen un histograma de intensidades más disperso en esta región comparado con el de los controles. Esto nos puede indicar una disminución en la captación del $[^{18}F]FDG$ en la región, o bien que se tiene una distribución no uniforme, ambas posibilidades podrían estar relacionadas con la sospecha de daño en el tálamo debido a la acumulación de placa $A\beta$.

En el tercer caso de estudio, las únicas características obtenidas del análisis de correlación que no forman parte del patrón típico de la EA son el tálamo derecho y la materia blanca, como se aprecia en la figura 3.8. El tálamo derecho, como se mencionó en la discusión del segundo caso de estudio, podría verse afectado en etapas tempranas de la enfermedad por las placas seniles. En cuanto a la materia blanca, a pesar de utilizarse frecuentemente como región de referencia en estudios $[^{18}F]FDG$ -PET, se ha demostrado que en esta zona se producen cambios en etapas previas y durante la EA, asociados a la acumulación de las placas $A\beta$ e incluso se han realizado estudios en los que se utilizan características radiómicas medidas en la materia blanca en estudios $[^{18}F]FDG$ -PET para predecir el progreso del MCI a la EA [64]. Las características con las que se obtuvieron mayores coeficientes de correlación de Spearman con el diagnóstico clínico se midieron en el GFS derecho y en el GPS derecho, las cuales como se mencionó se encuentran dentro del patrón hipometabólico típico de la EA. En el GFS la característica medida con mayor correlación fue la $IntensHist_{Curtosis}$, la cual al igual que en el segundo caso de estudio es una medida de qué tan angosta es la distribución de niveles de gris en la región, la única diferencia es que se discretizan los niveles de gris para generar un histograma de intensidades con un número fijo de bins. El comportamiento de la característica entre pacientes con sospecha de EA y controles es casi idéntico al que se presentó en el segundo caso de estudio, lo cual como se mencionó puede ser un indicativo de que se redujeron la cantidad de zonas con intensidad promedio en la región. En el GPS derecho, las características con mayor correlación de Spearman fueron la $NGTDM_{Tosquedad}$ y la $GLCM_{Correlacion}$. Ambas características también obtuvieron un mayor coeficiente de correlación de Pearson con la clasificación de PMOD en estas mismas regiones y como se explicó, la tosquedad nos indica la tasa de cambio espacial de intensidad y la correlación es una medida de uniformidad. Los diagramas de caja correspondientes a las características nos muestran, como se mencionó anteriormente, que los pacientes con sospecha de EA presentan una uniformidad y tasa de cambio o gradiente de captación del radiofármaco menores, lo cual es consistente con la EA.

En los tres casos de estudio, la mayoría de las características correlacionadas con la clasificación

de los pacientes fueron medidas en regiones del lado derecho, aunque también se encontraron algunas en el lado izquierdo. En cambio, en la figura 3.5 se observan más regiones anormales con respecto a la EA en el lado izquierdo del cerebro del paciente. Sin embargo, se debe recordar que muchas de las características con mayores correlaciones, tienen coeficiente negativo. Además, el ejemplo presentado en la figura 3.5 es de un solo paciente, por lo que es necesario un estudio específico acerca de la lateralidad de la enfermedad para poder inferir que el daño ocasionado por la EA se produce principalmente en alguno de los lados del cerebro.

En el análisis de Mann Whitney U, se observa que todas las características presentadas en los diagramas de caja obtuvieron valores p muy inferiores al valor de referencia típico de 0.05 [48], lo cual sugiere que los valores de las características medidas en ambos grupos de pacientes para los tres casos de estudio, pueden ser considerados diferentes. Esto nos indica que dichas características medidas en esas regiones podrían ser utilizadas como valores de referencia para asistir en el diagnóstico de la EA, sin embargo, es necesario realizar más estudios al respecto que demuestren que estos resultados son reproducibles con otros pacientes.

Los resultados de rendimiento de los diferentes algoritmos entrenados con todas las características indicadas en la figura 2.7, son variados. En el primer caso de estudio, se obtuvieron buenos resultados de EC para los tres algoritmos, siendo la red neuronal y el KNN los que obtuvieron los mejores resultados. También se puede observar que los algoritmos de KNN y SVM obtuvieron AUC aceptables que sugieren buena exactitud para algunos propósitos y el algoritmo de red neuronal obtuvo una AUC muy buena, que sugiere alta exactitud [51]. Los algoritmos KNN y SVM obtuvieron valores de especificidad perfecta, pero con baja sensibilidad, lo cual puede ocurrir debido a que la muestra para realizar las predicciones es de 15 pacientes (30% del total de los pacientes de este grupo). Para que un modelo sea un buen predictor, debe tener alta sensibilidad y especificidad. En el segundo caso de estudio, se observa que únicamente se obtuvieron resultados aceptables con el algoritmo tipo red neuronal, con una buena exactitud en la clasificación y una AUC muy buena. Con los algoritmos de KNN y SVM se obtuvieron EC bajas y AUC bajas, las cuales son un indicativo de baja exactitud. En el tercer caso de estudio, se observa que nuevamente se obtuvo una EC baja con los tres algoritmos, sin embargo, con el algoritmo de red neuronal se obtuvo una AUC aceptable y con el algoritmo Random Forest se obtuvo una AUC muy buena, por lo que utilizando todas las características de entrada en este caso de estudio, fue el único algoritmo con buena exactitud. Además, cabe destacar que el algoritmo Random Forest fue el que requirió menor tiempo de entrenamiento en todos los casos de estudio con 4.437 segundos, este tiempo es mucho menor que el tiempo que requirieron los demás algoritmos para entrenamiento. Este parámetro puede ser relevante al utilizar una gran cantidad de instancias y características en el entrenamiento de un algoritmo de aprendizaje autónomo.

Los resultados de rendimiento obtenidos utilizando las características seleccionadas con un coeficiente de correlación de Pearson mayor que ± 0.6 mejoraron de manera general, como se puede observar en la tabla 3.4. En el primer caso de estudio, se observa que los tres algoritmos con los mejores rendimientos obtuvieron resultados idénticos, únicamente con tiempos de entrenamiento diferentes. También se puede notar que con el algoritmo de red neuronal se obtuvieron resultados similares a las pruebas realizadas utilizando todas las características de entrada, sin embargo, los otros dos algoritmos con buenos resultados en este caso fueron el

Random Forest y el algoritmo de Regresión Logística. Con los tres algoritmos se obtuvo una AUC muy buena, lo que sugiere que los modelos tienen una exactitud alta. En el segundo caso de estudio, se puede ver que los resultados de rendimiento mejoraron considerablemente utilizando estas características y que de nueva cuenta, además del algoritmo de red neuronal, los otros dos algoritmos con mejores resultados fueron el Random Forest y Regresión Logística. Se obtuvieron mejores resultados en la EC con respecto al uso de todas las características, se tuvieron AUC muy buenas que reflejan exactitud alta por parte de los modelos y también hubo un buen balance entre sensibilidad y especificidad por parte de los tres algoritmos. En el tercer caso de estudio también se obtuvieron mejores resultados de rendimiento, sin embargo, los algoritmos con los mejores resultados fueron los mismos que los que se tuvieron utilizando todas las características de entrada. Se puede notar que con la red neuronal se obtuvo una EC idéntica, pero con los algoritmos de Random Forest y SVM este valor mejoró. Con los tres algoritmos se obtuvieron AUC muy buenas y en esta ocasión se tuvieron valores bajos de sensibilidad y especificidad. Algo importante a destacar, es que todos los algoritmos en los tres casos de estudio utilizando las características con mayores coeficientes de correlación de Pearson, redujeron significativamente sus tiempos de entrenamiento. Esto se vuelve aún más relevante cuando se trabaja con bases de datos que cuentan con grandes cantidades de instancias, ya que el tiempo de entrenamiento es mucho mayor y se requiere de procesos de depuración de características.

En la última selección de características, utilizando las que obtuvieron un coeficiente de correlación de Spearman mayor que ± 0.6 con la clasificación de los pacientes en cada caso de estudio, los resultados de rendimiento mejoraron con respecto a la selección de características mediante el coeficiente de correlación de Pearson de manera general, como se puede observar en la tabla 3.5. En el primer caso de estudio, nuevamente los algoritmos de red neuronal y Random Forest obtuvieron los mejores resultados, además del algoritmo de SVM. Los tres algoritmos obtuvieron una EC y AUC muy buenas, lo que sugiere que los tres modelos tienen una alta exactitud en sus clasificaciones. Además, también se tuvo un buen balance entre sensibilidad y especificidad, por lo que para este caso de estudio, la selección de características mediante la correlación de Spearman fue óptima. En el segundo caso de estudio, los algoritmos con los mejores resultados fueron el KNN, SVM y de nueva cuenta el Random Forest, se puede notar que este fue el único caso en el que la red neuronal no fue uno de los algoritmos con mejores resultados. En general, se obtuvieron resultados de rendimiento similares a los obtenidos con la selección de características de acuerdo con la correlación de Pearson, reduciendo incluso la EC en el caso del algoritmo Random Forest. Las AUC obtenidas mediante los tres algoritmos fueron muy buenas y también se tuvieron buenos valores de sensibilidad y especificidad. A pesar de la disminución en la EC obtenida mediante el algoritmo de Random Forest con esta selección de características, se debe destacar que las predicciones con este algoritmo pueden cambiar al volver a entrenarlo, obteniendo valores similares pero no iguales. Por lo tanto, para este caso de estudio, ambas selecciones de características con los métodos de correlación fueron óptimas. En el tercer caso de estudio, los algoritmos con mejores resultados fueron la red neuronal, SVM y KNN. Nuevamente, los resultados de rendimiento mejoraron con respecto a la selección de características con la correlación de Pearson, obteniendo muy buenos valores en la EC y en el AUC. También se tuvieron buenos valores de sensibilidad y especificidad por lo que, en este caso de estudio, la selección de características mediante la correlación de Spearman con la clasificación del software PMOD fue la óptima. Algo importante a destacar, es que para las pruebas realizadas con la selección de características con mayor coeficiente de correlación

de Spearman en los tres casos de estudio, el algoritmo que obtuvo los mejores resultados fue el de SVM, lo cual puede deberse a que este algoritmo brinda buenos resultados con muestras pequeñas [34], sumado a una selección adecuada de características.

En general, se obtuvieron resultados aceptables en la mayoría de las pruebas realizadas. Como se mencionó, hubo una mejoría considerable al realizar selección de características para el entrenamiento de los algoritmos de aprendizaje autónomo. En el primer y tercer caso de estudio, la selección de características con mayor coeficiente de correlación de Spearman con la clasificación de los pacientes fue con la que se obtuvieron los mejores resultados, mientras que en el segundo caso de estudio, los resultados realizando ambas selecciones estuvieron a la par, por lo que se puede afirmar que en este estudio la selección óptima de características se realizó con el método de correlación de Spearman. A pesar de haber comprobado que es posible obtener características con valores diferentes para cada tipo de paciente mediante los análisis de correlación y la prueba de Mann Whitney U, la implementación de los algoritmos de aprendizaje autónomo es necesaria ya que estos modelos no solo toman en cuenta los valores de una característica, sino el patrón de valores de características completo. Además, considerar solo una o unas cuantas medidas como referencia en el apoyo al diagnóstico de la EA no sería muy preciso debido a una característica del cerebro denominada plasticidad, la cual le permite cambiar sus funciones en respuesta a estímulos intrínsecos y extrínsecos, volviendo difícil el generalizar las funciones cerebrales de regiones específicas con todos los pacientes [65]. Sin embargo, al analizar el patrón completo de la enfermedad y seguir nutriendo a los algoritmos de aprendizaje autónomo con instancias (pacientes), es posible entenderla de mejor manera y brindar un apoyo al diagnóstico de la EA mediante las predicciones o probabilidades de clasificación. Particularmente, el algoritmo de tipo red neuronal es prometedor en ese sentido, ya que realiza correcciones en la toma de decisiones de las redes conforme se va nutriendo con más instancias.

Finalmente, analizando los resultados obtenidos mediante la clasificación por clústers con el algoritmo K-means, se puede destacar que se obtuvieron valores de EC aceptables y que este algoritmo tiene la ventaja de no requerir una división en datos de entrenamiento y prueba. Además, al ser un algoritmo de aprendizaje autónomo no supervisado, no requiere de una clase objetivo y únicamente clasifica a las instancias por los patrones de sus características. Con la EC obtenida en cada caso, se puede tener una idea de qué tan confiable sería la clasificación dada por este algoritmo como apoyo en el diagnóstico de la EA.

Capítulo 5

Conclusiones

La herramienta de análisis de pacientes con posible EA del software PMOD, permitió tener un parámetro de referencia en cuanto al tipo de pacientes que se tienen, ya que la clasificación dada por el software fue casi idéntica a la clasificación dada por el diagnóstico clínico, sin embargo, el análisis que realiza es limitado y puede verse afectado por muchos factores. Además, solo es efectivo utilizando pacientes con sospecha clínica de EA y controles, siendo incapaz de discriminar entre pacientes con alguna enfermedad neurológica y pacientes con sospecha de EA, por lo que es necesario crear e implementar algoritmos que permitan discernir entre ambos pacientes, ya que en algunos casos se pueden tener dos o más diagnósticos probables y un parámetro de referencia dado por un algoritmo podría ser un apoyo importante.

El análisis de correlación de Spearman, en conjunto con la prueba de Mann Whitney U, demostró ser efectivo para encontrar características medidas en diferentes regiones cerebrales que muestren un comportamiento distinto para cada clase de paciente ya que, mediante los diagramas de caja, fue posible observar la diferencia en la distribución de los valores de estas características en los pacientes con sospecha de EA y con alguna enfermedad neurológica o controles. Estas diferencias pudieron ser comprobadas cuantitativamente a través de la prueba de Mann Whitney U, eliminando la subjetividad del análisis visual. Estos resultados son importantes para establecer valores de referencia que permitan asistir el diagnóstico de la EA, sin embargo, es necesario realizar un mejor protocolo para la selección y clasificación de pacientes, además de una muestra mucho mayor, para poder plantear el establecimiento de dichos valores. Además como se mencionó, se tiene la limitante de la plasticidad neuronal, por lo que es necesario realizar un estudio multidisciplinario para discutir el establecimiento de medidas de referencia en regiones cerebrales para apoyar el diagnóstico de la EA.

Mediante las pruebas realizadas con los modelos de aprendizaje autónomo creados utilizando todas las características de entrada, se obtuvieron resultados aceptables con algunos algoritmos. A pesar de no obtenerse rendimientos tan altos como los obtenidos con la selección de características importantes, se tiene la ventaja de analizar el patrón cerebral completo de la enfermedad. Sin embargo, aunque no se tenía una muestra de pacientes extensa, los tiempos de entrenamiento de los algoritmos fueron considerablemente altos, por lo que en futuros estudios en los que se utilicen más instancias se debe considerar la selección de características importantes para evitar tiempos de entrenamiento muy largos.

La selección de características utilizando como referencia la correlación de Pearson permitió mejorar considerablemente los resultados de rendimiento de los diferentes algoritmos, por lo que no debe ser descartado su uso para la selección de características importantes. De manera global, los mejores resultados de rendimiento de los algoritmos se obtuvieron mediante la selección de características importantes considerando la correlación de Spearman, por lo tanto, es recomendable utilizar este método en futuros estudios, especialmente si se tienen muestras grandes de pacientes o bases de datos, ya que los tiempos de entrenamiento se redujeron considerablemente. Tomando en cuenta los mejores resultados de rendimiento, podemos afirmar que el uso de los algoritmos de aprendizaje autónomo, en conjunto con medidas radiómicas y protocolos clínicos adecuados, son herramientas factibles para apoyar en el diagnóstico de la EA y no se descarta que en un futuro puedan ser utilizadas como referencias clínicas.

Capítulo 6

Trabajo a futuro

En este trabajo se obtuvieron resultados prometedores, sin embargo, es necesario mejorar en algunos aspectos de las diferentes etapas. Lo primero que se debe mejorar en un trabajo a futuro, es el protocolo para la adquisición de los estudios de pacientes, dado que en este trabajo se adquirieron las imágenes de manera retrospectiva, se propone que para trabajos futuros se realice un protocolo completo que mantenga un control sobre la adquisición de los estudios PET y que para todos los pacientes se cuente con el resultado de algún test neuropsicológico aplicado previamente por su médico tratante, o bien, aplicarlo en la unidad. También es deseable que el trabajo sea multidisciplinario para contar con la opinión de médicos y psicólogos en la decisión del diagnóstico del paciente, de esta manera se puede tener mayor certeza al momento de establecer la clasificación real de los pacientes y esto brindaría resultados más confiables.

De igual forma se propone que para trabajos futuros, durante el preprocesamiento de las imágenes, se utilicen métodos para evitar diferencias en intensidad o resolución debidas a los equipos donde se adquirieron los estudios ya que, a pesar de reducir estos efectos en el proceso de medición de características radiómicas, estos se podrían hacer presentes si las diferencias entre las características de los equipos son grandes.

Otra propuesta a futuro es realizar el entrenamiento de los algoritmos de aprendizaje autónomo mediante el uso de bases de datos reconocidas como la de la Iniciativa de Neuroimagen de Enfermedad de Alzheimer (ADNI), esto permitiría entrenar de mejor manera a los algoritmos y utilizando métodos para evitar efectos por diferencia entre los equipos de adquisición de imagen, se podrían realizar predicciones más precisas acerca del diagnóstico de los pacientes.

También se propone considerar como característica de entrada el resultado (positivo o negativo) de estudios con radiofármacos específicos para la acumulación de placa $A\beta$ como el $[^{11}\text{C}]\text{Pittsburgh}$, esto ayudaría a tener una clasificación más precisa de los pacientes, descartando falsos positivos por EA.

Finalmente, de realizarse un mejor estudio, se propone plantear el uso de las probabilidades de clasificación de los algoritmos de aprendizaje autónomo como valores de referencia que permitan asistir en el diagnóstico de la Enfermedad de Alzheimer.

Bibliografía

- [1] E. Marieb, Anatomía y Fisiología Humana, 9th Edition, Pearson, 2008.
- [2] J. Molinuevo, J. Peña, Guía oficial para la práctica clínica en demencias: conceptos, criterios y recomendaciones, Sociedad Española de Neurología, 2009.
URL https://www.sen.es/pdf/guias/Guia_oficial_de_practica_clinica_en_demencias_2009.pdf
- [3] F. López, M. Reyna, Demencias: una visión panorámica, Universidad Autónoma de San Luis Potosí, 2015.
URL <http://www.geriatria.salud.gob.mx/descargas/publicaciones/Demencia.pdf>
- [4] González Hernández, Principios de bioquímica clínica y patología molecular, 3rd Edition, Elsevier, 2019.
- [5] Z. Chen, C. Zhong, Decoding alzheimer's disease from perturbed cerebral glucose metabolism: Implications for diagnostic and therapeutic strategies, Progress in Neurobiology 108 (2013) 21–43. doi:<https://doi.org/10.1016/j.pneurobio.2013.06.004>.
URL <https://www.sciencedirect.com/science/article/pii/S0301008213000531>
- [6] L. Schilling, E. Zimmer, M. Shin, A. Leuzy, T. Pascoal, A. Benedet, W. Borelli, A. Palmi, S. Gauthier, P. Rosa, Imaging Alzheimer's Disease pathophysiology with PET, Dement Neuropsychol 10 (2) (2016) 79–90. doi:[10.1590/S1980-5764-2016DN1002003](https://doi.org/10.1590/S1980-5764-2016DN1002003).
- [7] Y. Ding, J. H. Sohn, M. G. Kawczynski, H. Trivedi, R. Harnish, N. W. Jenkins, D. Lituev, T. P. Copeland, M. S. Aboian, C. Mari Aparici, S. C. Behr, R. R. Flavell, S.-Y. Huang, K. A. Zalocusky, L. Nardo, Y. Seo, R. A. Hawkins, M. Hernandez Pampaloni, D. Hadley, B. L. Franc, A Deep Learning Model to Predict a Diagnosis of Alzheimer Disease by Using 18F-FDG PET of the Brain, Radiology 290 (2) (2019) 456–464, PMID: 30398430. arXiv:<https://doi.org/10.1148/radiol.2018180958>, doi:[10.1148/radiol.2018180958](https://doi.org/10.1148/radiol.2018180958).
URL <https://doi.org/10.1148/radiol.2018180958>
- [8] P. Zhou, R. Zeng, L. Yu, Y. Feng, C. Chen, F. Li, Y. Liu, Y. Huang, Z. Huang, t. A. D. N. I. , Deep-Learning Radiomics for Discrimination Conversion of Alzheimer's Disease in Patients With Mild Cognitive Impairment: A Study Based on 18F-FDG PET Imaging, Frontiers in Aging Neuroscience 13 (2021). doi:[10.3389/fnagi.2021.764872](https://doi.org/10.3389/fnagi.2021.764872).
URL <https://www.frontiersin.org/article/10.3389/fnagi.2021.764872>

-
- [9] W. Bao, F. Xie, C. Zuo, Y. Guan, Y. H. Huang, Pet neuroimaging of alzheimer's disease: Radiotracers and their utility in clinical research, *Frontiers in Aging Neuroscience* 13 (2021). doi:10.3389/fnagi.2021.624330.
URL <https://www.frontiersin.org/articles/10.3389/fnagi.2021.624330>
- [10] S. Cherry, *Physics in nuclear medicine*, 4th Edition, Elsevier, 2012.
- [11] A. Chandra, P.-E. Valkimadi, G. Pagano, O. Cousins, G. Dervenoulas, M. Politis, for the Alzheimer's Disease Neuroimaging Initiative, Applications of amyloid, tau, and neuroinflammation PET imaging to Alzheimer's Disease and Mild Cognitive Impairment, *Human Brain Mapping* 40 (18) (2019) 5424–5442. arXiv:<https://onlinelibrary.wiley.com/doi/pdf/10.1002/hbm.24782>, doi:<https://doi.org/10.1002/hbm.24782>.
URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/hbm.24782>
- [12] A. R. Sowa, I. M. Jackson, T. J. Desmond, J. Alicea, A. J. Mufarreh, J. M. Pham, J. Stauff, W. P. Winton, M. V. Fawaz, B. D. Henderson, B. G. Hockley, V. E. Rogers, R. A. Koeppe, P. J. H. Scott, Futureproofing [18F]Fludeoxyglucose manufacture at an Academic Medical Center, *EJNMMI Radiopharmacy and Chemistry* 3 (1) (2018) 12. doi:10.1186/s41181-018-0048-x.
URL <https://doi.org/10.1186/s41181-018-0048-x>
- [13] E. Segtnan, S. Hess, P. Grupe, P. Hoiland, F18-fluorodeoxyglucose PET/Computed tomography for primary brain tumors , *PET Clinics* 10 (1) (2015) 59–73.
- [14] A. Glaudemans, R. Enting, M. Heesters, et al, Value of 11C-methionine PET in imaging brain tumours and metastases , *European Journal of Nuclear Medicine and Molecular Imaging* 40 (2013) 615–635.
- [15] A. Prem Kumar, N. Singh, D. Nair, A. Justin, Neuronal PET tracers for Alzheimer's Disease, *Biochemical and Biophysical Research Communications* 587 (2022) 58–62. doi:<https://doi.org/10.1016/j.bbrc.2021.11.093>.
URL <https://www.sciencedirect.com/science/article/pii/S0006291X21016181>
- [16] G. Martínez, R. Vernooij, P. Fuentes, P. Zamora, J. Bonfill, L. Flicker, 18F PET with florbetapir for the early diagnosis of Alzheimer's disease dementia and other dementias in people with mild cognitive impairment (MCI), *Cochrane Database of Systematic Reviews* (11) (2017). doi:10.1002/14651858.CD012216.pub2.
URL <https://doi.org/10.1002/14651858.CD012216.pub2>
- [17] D. Bailey, J. Humm, A. Todd-Pokropek, A. v. Aswegen, *Nuclear medicine physics: a handbook for teachers and students*, 1st Edition, International Atomic Energy Agency, 2014.
- [18] P. Saxena, A. Singh, M. Dixit, S. Kheruka, T. Mahmood, S. Gambhir, Establishing the [18F]-FDG production via two different automated synthesizers for routine clinical studies: Our institutional experiences of 4 years, *Indian Journal of Nuclear Medicine* 36 (2) (2021) 120–124. doi:10.4103/ijnm.IJNM_137_20.
-

-
- [19] K. Kawada, M. Iwamoto, Y. Sakai, Mechanisms underlying 18f-fluorodeoxyglucose accumulation in colorectal cancer, *World Journal of Radiology* 8 (10) (2016) 880–886. doi:10.4329/wjr.v8.i11.880.
- [20] R. K. J. Brown, N. I. Bohnen, K. K. Wong, S. Minoshima, K. A. Frey, Brain PET in Suspected Dementia: Patterns of Altered FDG Metabolism, *RadioGraphics* 34 (3) (2014) 684–701, PMID: 24819789. arXiv:https://doi.org/10.1148/rg.343135065, doi:10.1148/rg.343135065.
URL https://doi.org/10.1148/rg.343135065
- [21] E. Guedj, A. Varrone, R. Boellaard, EANM procedure guidelines for brain PET imaging using [18F]FDG, version 3, *European Journal of Nuclear Medicine and Molecular Imaging* 49 (2022) 632–651. doi:https://doi.org/10.1007/s00259-021-05603-w.
- [22] A. Lammertsma, S. Hume, Simplified Reference Tissue Model for PET Receptor Studies, *NeuroImage* 4 (3) (1996) 153–158. doi:doi:10.1006/nimg.1996.0066.
- [23] I. L. Alves, A. T. Willemsen, R. A. Dierckx, A. M. M. da Silva, M. Koole, Dual time-point imaging for post-dose binding potential estimation applied to a [11C]raclopride PET dose occupancy study, *Journal of Cerebral Blood Flow & Metabolism* 37 (3) (2017) 866–876. doi:10.1177/0271678X16644463.
URL https://doi.org/10.1177/0271678X16644463
- [24] C. S. Patlak, R. G. Blasberg, J. D. Fenstermacher, Graphical evaluation of blood-to-brain transfer constants from multiple-time uptake data, *Journal of Cerebral Blood Flow & Metabolism* 3 (1) (1983) 1–7, PMID: 6822610. arXiv:https://doi.org/10.1038/jcbfm.1983.1, doi:10.1038/jcbfm.1983.1.
URL https://doi.org/10.1038/jcbfm.1983.1
- [25] Èlia Torrent, M. Farré, I. Abasolo, O. Millan, J. Llop, J. D. Gispert, A. Ruiz, D. Pareto, Optimization of [11C]Raclopride Positron Emission Tomographic Rat Studies: Comparison of Methods for Image Quantification, *Molecular Imaging* 12 (4) (2013) 1157–1164. doi:10.2310/7290.2012.00040.
URL https://doi.org/10.2310/7290.2012.00040
- [26] P. E. Kinahan, J. W. Fletcher, Positron Emission Tomography-Computed Tomography Standardized Uptake Values in Clinical Practice and Assessing Response to Therapy, *Seminars in Ultrasound, CT and MRI* 31 (6) (2010) 496–505, pET-CT in RT Planning and Assessment. doi:https://doi.org/10.1053/j.sult.2010.10.001.
URL https://www.sciencedirect.com/science/article/pii/S0887217110000880
- [27] K. Park, S. Kim, What Do We Measure in Oncology PET? , *Nuclear Medicine and Molecular Imaging* 51 (2017) 212–216. doi:https://doi.org/10.1007/s13139-016-0416-y.
- [28] A. Sher, F. Lacoeyille, P. Fosse, et al., For avid glucose tumors, the SUV peak is the most reliable parameter for [18F]FDG-PET/CT quantification, regardless of acquisition time, *EJNMMI Research* 6 (21) (2016). doi:10.1186/s13550-016-0177-8.
-

-
- [29] R. Gillies, P. Kinahan, H. Hricak, Radiomics: Images are more than pictures, they are data, *Radiology* 278 (2) (2016) 563–577. doi:10.1097/MNM.0b013e328351d4f5.
- [30] J. Song, Y. Yin, H. Wang, Z. Chang, Z. Liu, L. Cui, A review of original articles published in the emerging field of radiomics, *European Journal of Radiology* 127 (2020) 108991. doi:https://doi.org/10.1016/j.ejrad.2020.108991.
URL <https://www.sciencedirect.com/science/article/pii/S0720048X20301807>
- [31] H. Clifton, A. Vial, A. Miller, C. Ritz, M. Field, L. Holloway, M. Ros, M. Carolan, D. Stirling, Using Machine Learning Applied to Radiomic Image Features for Segmenting Tumour Structures (2019) 1981–1988doi:10.1109/APSIPAASC47483.2019.9023077.
- [32] N. Otsu, A threshold selection method from gray-level histograms, *IEEE Transactions on Systems, Man, and Cybernetics* 9 (1) (1979) 62–66. doi:10.1109/TSMC.1979.4310076.
- [33] Texture User Guide. Local Imaging Features Extraction.
URL <https://www.lifexsoft.org/index.php/resources/documentation>
- [34] A. Geron, Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow, 2nd Edition, O'Reilly Media, Inc, 2019.
- [35] R. C. Deo, Machine Learning in Medicine, *Circulation* 132 (20) (2015) 1920–1930. doi:10.1161/CIRCULATIONAHA.115.001593.
- [36] X. Wu, V. Kumar (Eds.), *The Top Ten Algorithms in Data Mining*, 1st Edition, Chapman and Hall/CRC, 2009. doi:10.1201/9781420089653.
- [37] H. Aerts, E. Velazquez, R. Leijenaar, et al, Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach, *Nature* 5 (4006) (2014) 169–186. doi:https://doi.org/10.1038/ncomms5006.
- [38] C. Chang, X. Sun, G. Wang, H. Yu, W. Zhao, Y. Ge, S. Duan, X. Qian, R. Wang, B. Lei, L. Wang, L. Liu, M. Ruan, H. Yan, C. Liu, J. Chen, W. Xie, A Machine Learning Model Based on PET/CT Radiomics and Clinical Characteristics Predicts ALK Rearrangement Status in Lung Adenocarcinoma, *Frontiers in Oncology* 11 (2021). doi:10.3389/fonc.2021.603882.
URL <https://www.frontiersin.org/article/10.3389/fonc.2021.603882>
- [39] J. Jiang, J. Zhang, Z. Li, L. Li, B. Huang, A. D. N. I. , Using Deep Learning Radiomics to Distinguish Cognitively Normal Adults at Risk of Alzheimer's Disease From Normal Control: An Exploratory Study Based on Structural MRI, *Frontiers in Medicine* 9 (2022). doi:10.3389/fmed.2022.894726.
URL <https://www.frontiersin.org/article/10.3389/fmed.2022.894726>
- [40] P. Alongi, R. Laudicella, F. Panasiti, A. Stefano, A. Comelli, P. Giaccone, A. Arnone, F. Minutoli, N. Quartuccio, C. Cupidi, G. Arnone, T. Piccoli, L. M. E. Grimaldi, S. Baldari, G. Russo, Radiomics Analysis of Brain [18F]FDG PET/CT to Predict Alzheimer's Disease in Patients with Amyloid PET Positivity: A Preliminary Report on the Application of SPM Cortical Segmentation, Pyradiomics and Machine-Learning Analysis, *Diagnostics*
-

- 12 (4) (2022). doi:10.3390/diagnostics12040933.
URL <https://www.mdpi.com/2075-4418/12/4/933>
- [41] Y. Li, J. Jiang, J. Lu, J. Jiang, H. Zhang, C. Zuo, Radiomics: a novel feature extraction method for brain neuron degeneration disease using 18F-FDG PET imaging and its implementation for Alzheimer's Disease and Mild Cognitive Impairment, *Therapeutic Advances in Neurological Disorders* 12 (2019) 1756286419838682, PMID: 30956687. arXiv: <https://doi.org/10.1177/1756286419838682>, doi:10.1177/1756286419838682. URL <https://doi.org/10.1177/1756286419838682>
- [42] PMOD Technologies LLC, PMOD - The PET/MR and PET/CT Quantification and Analysis Software, <https://www.pmod.com>, consultado el 17 de marzo de 2023 (2023).
- [43] K. Herholz, E. Salmon, D. Perani, J. Baron, V. Holthoff, L. Frölich, P. Schönknecht, K. Ito, R. Mielke, E. Kalbe, et al., Discrimination between Alzheimer dementia and controls by automated analysis of multicenter FDG PET, *NeuroImage* 17 (1) (2002) 302–316. doi: <https://doi.org/10.1006/nimg.2002.1208>.
- [44] D. Wayne, *Bioestadística: Base para el análisis de las ciencias de la salud*, 4th Edition, LIMUSA Wiley, México, D.F., 2016.
- [45] K. Herholz, S. Westwood, C. Haense, G. Dunn, Evaluation of a Calibrated 18F-FDG PET Score as a Biomarker for Progression in Alzheimer Disease and Mild Cognitive Impairment, *Journal of Nuclear Medicine* 52 (8) (2011) 1218–1226. doi:10.2967/jnumed.111.090902.
- [46] A. Hammers, R. Allom, M. J. Koeppe, S. L. Free, R. Myers, L. Lemieux, T. N. Mitchell, D. J. Brooks, J. S. Duncan, Three-dimensional maximum probability atlas of the human brain, with particular reference to the temporal lobe, *Human Brain Mapping* 19 (4) (2003) 224–247. arXiv:<https://onlinelibrary.wiley.com/doi/pdf/10.1002/hbm.10123>, doi:<https://doi.org/10.1002/hbm.10123>. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/hbm.10123>
- [47] A. Zwanenburg, M. Vallières, M. A. Abdalah, H. J. W. L. Aerts, V. Andrearczyk, A. Apte, S. Ashrafinia, S. Bakas, R. J. Beukinga, R. Boellaard, et al., The Image Biomarker Standardization Initiative: Standardized Quantitative Radiomics for High-Throughput Image-based Phenotyping, *Radiology* 295 (2) (2020) 328–338. doi:10.1148/radiol.2020191145. URL <http://dx.doi.org/10.1148/radiol.2020191145>
- [48] G. W. Corder, *Nonparametric Statistics: A Step-by-Step Approach*, 2nd Edition, Wiley, Hoboken, NJ, 2014.
- [49] H. W. Lilliefors, On the kolmogorov-smirnov test for the exponential distribution, *Journal of the American Statistical Association* 62 (318) (1967) 399–402.
- [50] S. S. Shapiro, M. B. Wilk, An analysis of variance test for normality (complete samples), *Biometrika* 52 (3-4) (1965) 591–611.

-
- [51] J. A. Swets, Measuring the accuracy of diagnostic systems, *Science* 240 (4857) (1988) 1285–1293.
URL <http://www.jstor.org/stable/1701052>
- [52] R. Trevethan, Sensitivity, specificity, and predictive values: Foundations, pliabilitys, and pitfalls in research and practice, *Frontiers in Public Health* 5 (2017). doi:10.3389/fpubh.2017.00307.
URL <https://www.frontiersin.org/articles/10.3389/fpubh.2017.00307>
- [53] M. Jenkinson, C. F. Beckmann, T. E. Behrens, M. W. Woolrich, S. M. Smith, Fsl, *NeuroImage* 62 (2) (2012) 782–790, 20 YEARS OF fMRI. doi:<https://doi.org/10.1016/j.neuroimage.2011.09.015>.
URL <https://www.sciencedirect.com/science/article/pii/S1053811911010603>
- [54] B. Mena Montes, S. Torres Castro, T. Álvarez Cisneros, R. Rabaneda Bueno, N. M. Torres Carrillo, B. A. Jiménez Gallegos, Factores de protección de las demencias: importancia para las mujeres, *Género y salud en cifras* 15 (3) (2017) 12–17.
URL https://www.gob.mx/cms/uploads/attachment/file/315224/revista_15-3-1.pdf
- [55] C. Haense, K. Herholz, W. Jagust, W. Heiss, Performance of FDG PET for detection of Alzheimer’s disease in two independent multicentre samples (NEST-DD and AD-NI), *Dementia and Geriatric Cognitive Disorders* 28 (3) (2009) 259–266. doi:<https://doi.org/10.1159/000241879>.
- [56] J. H. Lee, J. Ryan, C. Andreescu, H. Aizenstein, H. K. Lim, Brainstem morphological changes in alzheimer’s disease, *NeuroReport* 26 (7) (2015) 411–415. doi:10.1097/WNR.0000000000000362.
- [57] D. J. Braun, L. J. Van Eldik, In vivo brainstem imaging in alzheimer’s disease: Potential for biomarker development, *Frontiers in Aging Neuroscience* 10 (2018). doi:10.3389/fnagi.2018.00266.
URL <https://www.frontiersin.org/articles/10.3389/fnagi.2018.00266>
- [58] E. Hoxha, P. Lippiello, F. Zurlo, I. Balbo, R. Santamaria, F. Tempia, M. C. Miniaci, The Emerging Role of Altered Cerebellar Synaptic Processing in Alzheimer’s Disease, *Frontiers in Aging Neuroscience* 10 (2018) 396. doi:<https://doi.org/10.3389/fnagi.2018.00396>.
- [59] H. I. L. Jacobs, D. A. Hopkins, H. C. Mayrhofer, E. Bruner, F. W. van Leeuwen, W. Raaijmakers, J. D. Schmahmann, The cerebellum in Alzheimer’s Disease: evaluating its role in cognitive decline, *Brain* 141 (1) (2018) 37–47.
- [60] J. P. Aggleton, A. Pralus, A. J. D. Nelson, M. Hornberger, Thalamic pathology and memory loss in early Alzheimer’s Disease: moving the focus from the medial temporal lobe to Papez circuit, *Brain* 139 (7) (2016) 1877–1890.
-

-
- [61] S. P. Poulin, R. Dautoff, J. C. Morris, L. F. Barrett, B. C. Dickerson, Amygdala atrophy is prominent in early Alzheimer's Disease and relates to symptom severity, *Psychiatry Research: Neuroimaging* 194 (1) (2011) 7–13. doi:<https://doi.org/10.1016/j.psychresns.2011.06.014>.
URL <https://www.sciencedirect.com/science/article/pii/S092549271100237X>
- [62] J. M. Burns, J. E. Galvin, C. M. Roe, J. C. Morris, D. W. McKeel, The pathology of the substantia nigra in alzheimer disease with extrapyramidal signs, *Neurology* 64 (8) (2005) 1397–1403. arXiv:<https://n.neurology.org/content/64/8/1397.full.pdf>, doi:10.1212/01.WNL.0000158423.05224.7F.
URL <https://n.neurology.org/content/64/8/1397>
- [63] Q. Yuan, X. Liang, C. Xue, W. Qi, S. Chen, Y. Song, H. Wu, X. Zhang, C. Xiao, J. Chen, Altered anterior cingulate cortex subregional connectivity associated with cognitions for distinguishing the spectrum of pre-clinical Alzheimer's Disease, *Frontiers in Aging Neuroscience* 14 (2022). doi:<https://doi.org/10.3389/fnagi.2022.1035746>.
- [64] J. Peng, W. Wang, Q. Song, J. Hou, H. Jin, X. Qin, Z. Yuan, Y. Wei, Z. Shu, 18f-fdg-pet radiomics based on white matter predicts the progression of mild cognitive impairment to alzheimer disease: A machine learning study, *Academic Radiology* (2022). doi:<https://doi.org/10.1016/j.acra.2022.12.033>.
URL <https://www.sciencedirect.com/science/article/pii/S1076633222006894>
- [65] P. Mateos-Aparicio, A. Rodríguez-Moreno, The impact of studying brain plasticity, *Frontiers in Cellular Neuroscience* 13 (2019). doi:10.3389/fncel.2019.00066.
URL <https://www.frontiersin.org/articles/10.3389/fncel.2019.00066>

Anexos

Definiciones y fundamento matemático

En estos anexos se definirán únicamente las características radiómicas más relevantes que se pueden obtener con el software empleado para este trabajo.

GLCM

La matriz de co-ocurrencia del nivel de gris es un arreglo de indicadores que toma en cuenta la co-ocurrencia de niveles de gris en pares de vóxeles con la finalidad de calcular características de textura [33]. La GLCM se calcula para 4 direcciones diferentes en 2D y para 13 direcciones diferentes en 3D con los vóxeles que se encuentran a una distancia (\vec{d}), como se muestra en la figura 4.1.

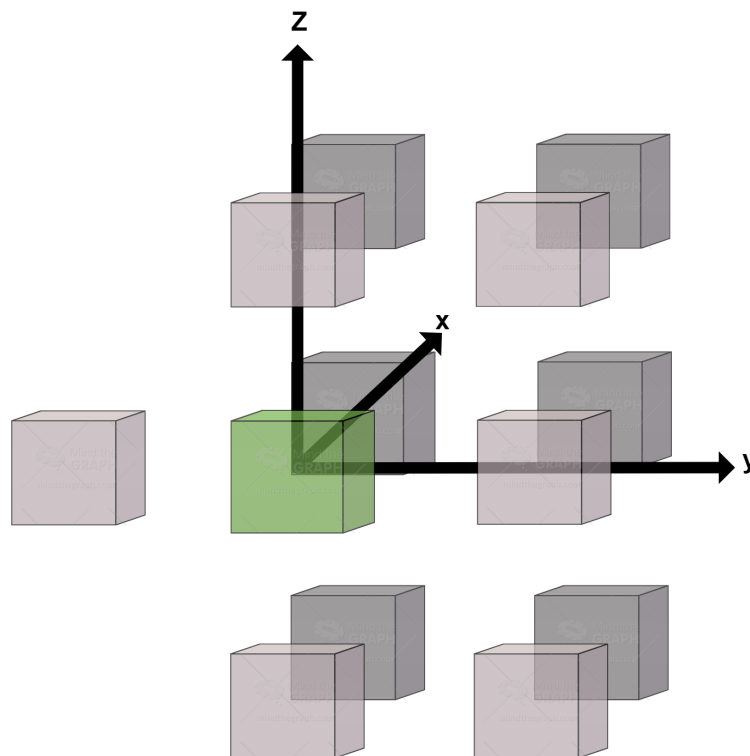


Figura 6.1: Direcciones posibles en 3D para la GLCM.

Cada índice representa el promedio de la co-ocurrencia de un nivel de gris para las 13 direcciones. Un elemento (i, j) de la matriz GLCM en una dirección en específico, se calcula de la siguiente manera

$$GLCM_{\Delta x, \Delta y}(i, j) = \frac{1}{\text{Pares}_{ROI}} \sum_{p=1}^{N-\Delta x} \sum_{q=1}^{M-\Delta y} \begin{cases} 1 & \text{si } (I(p, q) = i, I(p + \Delta x, q + \Delta y) = j) \\ & \& I(p, q), I(p + \Delta x, q + \Delta y) \in ROI \\ 0 & \text{de otra forma} \end{cases} \quad (6.1)$$

En donde $I(p, q)$ corresponde al vóxel con coordenadas (p, q) en una imagen I de tamaño $N \times M$, $\vec{d} = (\Delta x, \Delta y)$ corresponde a la distancia y cubre las 13 direcciones y Pares_{ROI} corresponde al total de pares de vóxeles en la región de interés. Como se mencionó, la GLCM describe la co-ocurrencia de niveles de gris de vóxeles vecinos que se encuentran separados una distancia δ , sin embargo, normalmente esta distancia se mantiene con un valor de 1, por lo que solo se calcula la co-ocurrencia de los vóxeles que rodean al vóxel $I(p, q)$. Del cálculo de la GLCM se derivan siete índices de textura que se definirán a continuación.

GLCM_Homogeneidad:

Este índice describe la homogeneidad de los pares de vóxeles de nivel de gris y se calcula como sigue

$$GLCM_Homogeneidad = \text{Promedio sobre 13(o 4) direcciones} \left(\sum_i \sum_j \frac{GLCM(i, j)}{1 + |i - j|} \right) \quad (6.2)$$

como podemos observar, mide que tan parecida es la matriz GLCM a una matriz identidad, es decir, la homogeneidad de la GLCM.

GLCM_Energía:

También llamada Uniformidad o Segundo Momento Angular, es la uniformidad de los pares de vóxeles de nivel de gris y se calcula de la siguiente manera

$$GLCM_Energía = \text{Promedio sobre 13(o 4) direcciones} \left(\sum_i \sum_j GLCM(i, j)^2 \right) \quad (6.3)$$

este parámetro permite comparar qué tanto varía la matriz de co-ocurrencia ya que al elevar al cuadrado cada elemento, se vuelve más notoria la diferencia entre las co-ocurrencias de los pares de vóxeles.

GLCM_Contraste: También llamada Varianza o Inercia, describe las variaciones locales en la GLCM y se calcula de la siguiente manera

$$GLCM_Contraste = \text{Promedio sobre 13(o 4) direcciones} \left(\sum_i \sum_j (i - j)^2 \cdot GLCM(i, j) \right) \quad (6.4)$$

al igual que la energía, permite evaluar variaciones de los niveles de gris de los vóxeles con respecto a sus vecinos, es decir, el contraste entre los diferentes niveles de gris.

GLCM_Correlación:

Es la dependencia lineal de los niveles de gris en la GLCM y se calcula mediante la siguiente expresión

$$\text{GLCM_Correlación} = \text{Promedio sobre 13(o 4) direcciones} \left(\frac{\sum_i \sum_j (i - \mu_i) \cdot (j - \mu_j) \cdot \text{GLCM}(i, j)}{\sigma_i \cdot \sigma_j} \right) \quad (6.5)$$

en donde μ_i o μ_j corresponden a los promedios en la fila i o columna j y σ_i o σ_j corresponden a las varianzas. Esta característica permite evaluar si existe dependencia lineal entre los pares de vóxeles de nivel de gris y que tan lineal es dicha dependencia.

GLCM_Entropía_log10:

Es la aleatoriedad de los pares de vóxeles de nivel gris y se mide mediante la siguiente expresión

$$\text{GLCM_Entropía}_{\log 10} = \left(- \sum_i \sum_j \text{GLCM}(i, j) \cdot \log_{10}(\text{GLCM}(i, j) + \varepsilon) \right) \quad (6.6)$$

en donde $\varepsilon = 2e^{-16}$. Como se mencionó, mide el grado de aleatoriedad de los vóxeles, sin embargo, cabe aclarar que para esta se utiliza la función log en base 10. La función logaritmo es útil para este fin ya que también se utiliza en física estadística para medir el número de microestados posibles de un sistema.

GLCM_Entropía_log2:

También mide la aleatoriedad de los pares de vóxeles de nivel de gris mediante la siguiente fórmula

$$\text{GLCM_Entropía}_{\log 2} = \left(- \sum_i \sum_j \text{GLCM}(i, j) \cdot \log_2(\text{GLCM}(i, j) + \varepsilon) \right) \quad (6.7)$$

pero como podemos observar, en este caso se utiliza la función logaritmo en base 2.

GLCM_Disimilitud:

Mide la variación de los pares de vóxeles de nivel de gris mediante la siguiente expresión

$$\text{GLCM_Disimilitud} = \left(\sum_i \sum_j |i - j| \cdot \text{GLCM}(i, j) \right) \quad (6.8)$$

podemos observar que la fórmula es muy parecida a la de contraste, pero en la disimilitud son menos notorias las diferencias locales.

GLCM_NID:

La diferencia inversa normalizada (NID) es una medida de homogeneidad. Mientras mayor sea la diferencia entre los niveles de gris de la matriz GLCM, menor será el valor de la NID y viceversa. Se calcula mediante la siguiente expresión

$$\text{GLCM}_{\text{NID}} = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \frac{p_{ij}}{1 + |i - j|/N_g} \quad (6.9)$$

en donde p_{ij} es la probabilidad conjunta de que los niveles de gris i y j ocurran en vóxeles vecinos y N_g es el número de niveles de gris discretizados presentes en la ROI.

NGLDM

La matriz de diferencia de nivel de gris del vecindario o NGLDM corresponde a la diferencia de niveles de gris entre un vóxel y sus 26 vecinos en 3 dimensiones u 8 en 2D. Un elemento $(i, 1)$ de la NGLDM corresponde a la probabilidad de ocurrencia del nivel i y el elemento $(i, 2)$ será igual a

$$NGLDM(i, 2) = \sum_p \sum_q \begin{cases} |\bar{M}(p, q) - i| & \text{si } I(p, q) = i \\ 0 & \text{de otra forma} \end{cases} \quad (6.10)$$

en donde $\bar{\mu}(p, q)$ es el promedio de intensidades sobre los 26 vóxeles vecinos del vóxel (p, q) . Se pueden calcular tres índices de textura a partir de la matriz NGLDM y se definen a continuación.

NGLDM_Tosquedad:

Es el nivel de la tasa de cambio espacial en la intensidad y se mide de la siguiente manera

$$NGLDM_Tosquedad = \frac{1}{\sum_i NGLDM(i, 1) \cdot NGLDM(i, 2)} \quad (6.11)$$

NGLDM_Contraste:

Es la diferencia de intensidad entre regiones vecinas y se calcula mediante la siguiente fórmula

$$NGLDM_Contraste = \left[\sum_i \sum_j NGLDM(i, 1) \cdot NGLDM(j, 1) \cdot (i - j)^2 \right] \cdot \frac{\sum_i NGLDM(i, 2)}{E \cdot G \cdot (G - 1)} \quad (6.12)$$

En donde E corresponde al número de vóxeles en el Volumen de Interés y G es el número de niveles de gris.

NGLDM_Ocupación:

Es la frecuencia espacial de los cambios de intensidad y se calcula de la siguiente manera

$$NGLDM_Ocupación = \frac{\sum_i NGLDM(i, 1) \cdot NGLDM(i, 2)}{\sum_i \sum_j |(i - NGLDM(i, 1) - j - NGLDM(j, 1))|} \quad (6.13)$$

con $NGLDM(i, 1) \neq 0, NGLDM(j, 1) \neq 0$

NGTDM

La matriz de diferencia de tono de gris del vecindario (NGTDM) contiene la suma de las diferencias de niveles de gris de los píxeles/vóxeles con nivel de gris discretizado i y el nivel de gris discretizado promedio de los píxeles/vóxeles vecinos dentro de una distancia Chebyshev δ . Básicamente es una alternativa a la GLCM, se calcula de manera similar pero considerando un tamaño de vecindario dado por la longitud de Chebyshev al momento de calcular la diferencia de niveles de gris discretizados.

GLRLM

La matriz de longitud de ejecución de nivel gris o GLLRM permite obtener el tamaño de las ejecuciones homogéneas para cada nivel de gris. Al igual que las otras matrices, la GLRLM se calcula para cada una de las 13 direcciones (o 4 en 2D) y cada índice que se obtiene de la matriz, representa en promedio en las 13 direcciones. Un elemento (i, j) de la GLRLM corresponde al

número de cadenas homogéneas de j vóxeles con intensidad i en una imagen, es decir, que las filas representan el nivel de gris y las columnas representan la longitud de una cadena, para una dirección en específico. De la GLRLM se derivan 11 índices que se describen a continuación.

GLRLM_SRE - GLRLM_LRE:

El Énfasis a corto plazo y el Énfasis a largo plazo describen la distribución de cadenas homogéneas cortas o largas respectivamente y se calculan de la siguiente manera

$$\text{GLRLM_SRE} = \text{Promedio sobre 13 (o 4) direcciones} \left(\frac{1}{H} \sum_i \sum_j \frac{\text{GLRLM}(i, j)}{j^2} \right) \quad (6.14)$$

$$\text{GLRLM_LRE} = \text{Promedio sobre 13 (o 4) direcciones} \left(\frac{1}{H} \sum_i \sum_j \text{GLRLM}(i, j) \cdot j^2 \right) \quad (6.15)$$

En donde H corresponde al número de cadenas homogéneas en el volumen de interés. Podemos notar que para el Énfasis a corto plazo, se hacen más evidentes las cadenas cortas al ser inversamente proporcional a la longitud y en el Énfases de cadena larga se vuelven más notorias las cadenas largas al ser directamente proporcional a la longitud.

GLRLM_LGRE - GLRLM_HGRE:

El énfasis de cadenas de nivel de gris bajo y el énfasis de cadenas de nivel de gris alto describen la distribución de las cadenas de nivel de gris bajo y alto respectivamente, mediante las siguientes expresiones

$$\text{GLRLM_LGRE} = \left(\frac{1}{H} \sum_i \sum_j \frac{\text{GLRLM}(i, j)}{i^2} \right) \quad (6.16)$$

$$\text{GLRLM_HGRE} = \left(\frac{1}{H} \sum_i \sum_j \text{GLRLM}(i, j) \cdot i^2 \right) \quad (6.17)$$

recordando que se calculan los promedios en las 13 direcciones. En este caso, observamos que el énfasis se le da a los niveles de gris bajos o altos que poseen las H cadenas de vóxeles.

GLRLM_SRLGE - GLRLM_SRHGE:

Énfasis de nivel de gris bajo en cadena corta y Énfasis de nivel de gris alto en cadena corta, como su nombre lo indica, son la distribución de cadenas homogéneas cortas con niveles de gris bajos y altos, respectivamente. Se pueden calcular de la siguiente manera

$$\text{GLRLM_SRLGE} = \left(\frac{1}{H} \sum_i \sum_j \frac{\text{GLRLM}(i, j)}{i^2 \cdot j^2} \right) \quad (6.18)$$

$$\text{GLRLM_SRHGE} = \text{Average over} \left(\frac{1}{H} \sum_i \sum_j \frac{\text{GLRLM}(i, j) \cdot i^2}{j^2} \right) \quad (6.19)$$

como podemos notar, estos índices hacen más evidentes las cadenas cortas que tienen niveles de gris bajos y altos, respectivamente.

GLRLM_LRLGE - GLRLM_LRHGE:

Énfasis de nivel de gris bajo de cadena larga y Énfasis de nivel de gris alto de cadena larga son lo contrario a los dos índices anteriores ya que describen la distribución de las cadenas largas de vóxeles homogéneos con niveles de gris bajos y altos respectivamente, mediante las siguientes expresiones

$$GLRLM_LRLGE = \left(\frac{1}{H} \sum_i \sum_j \frac{GLRLM(i, j) \cdot j^2}{i^2} \right) \quad (6.20)$$

$$GLRLM_LRHGE = \left(\frac{1}{H} \sum_i \sum_j GLRLM(i, j) \cdot i^2 \cdot j^2 \right) \quad (6.21)$$

podemos notar también en sus expresiones que de manera contraria a los dos índices anteriores, estos hacen énfasis en las cadenas largas de vóxeles homogéneos, para los niveles de gris bajos y altos.

GLRLM_GLNUR - GLRLM_RLNU:

La No-uniformidad de nivel de gris para corrida y la No-uniformidad de longitud de corrida describen la falta de uniformidad de los niveles de gris y la longitud de las corridas homogéneas y se calculan de la siguiente manera

$$GLRLM_GLNUR = \left(\frac{1}{H} \sum_i \left(\sum_j GLRLM(i, j) \right)^2 \right) \quad (6.22)$$

$$GLRLM_RLNU = \left(\frac{1}{H} \sum_j \left(\sum_i GLRLM(i, j) \right)^2 \right) \quad (6.23)$$

la primera medida hace evidente la falta de uniformidad de los niveles de gris al elevar al cuadrado la sumatoria correspondiente a los niveles de gris. La segunda medida permite notar la falta de uniformidad en el tamaño de las cadenas de vóxeles homogéneos.

GLRLM_RP:

El porcentaje de ejecución de la GLRLM mide la homogeneidad de las ejecuciones homogéneas mediante la siguiente fórmula

$$GLRLM_RP = \left(\frac{H}{\sum_i \sum_j (j \cdot GLRLM(i, j))} \right) \quad (6.24)$$

GLZLM

La matriz de longitud de zona de nivel de gris o GLZLM proporciona información sobre el tamaño de zonas homogéneas para cada nivel de gris en 3 dimensiones (o 2D). El elemento (i, j) de la GLZLM corresponde al número de zonas homogéneas de j vóxeles con una intensidad i en una imagen y se denomina $GLZLM(i, j)$. A partir de esta matriz se calculan 11 índices de textura que se describen a continuación.

GLZLM_SZE - GLZLM_LZE:

El Énfasis de zona corta y el Énfasis de zona larga describen la distribución de las zonas homogéneas cortas y largas respectivamente y se calculan de la siguiente manera

$$GLZLM_SZE = \frac{1}{H} \sum_i \sum_j \frac{GLZLM(i, j)}{j^2} \quad (6.25)$$

$$GLZLM_LZE = \frac{1}{H} \sum_i \sum_j GLZLM(i, j) \cdot j^2 \quad (6.26)$$

en donde H representa el número de zonas homogéneas en el volumen de interés. Podemos notar que al igual que en índices de otras matrices, se divide o se multiplica por el cuadrado del parámetro, en este caso el tamaño de la zona o número de vóxeles, para identificar de mejor manera las zonas cortas y largas.

GLZLM_LGZE - GLZLM_HGZE:

El Énfasis de zona de nivel de gris bajo y el Énfasis de zona de nivel de gris alto describen la distribución de las zonas de nivel de gris bajo y alto respectivamente y se calculan de la siguiente manera

$$GLZLM_LGZE = \frac{1}{H} \sum_i \sum_j \frac{GLZLM(i, j)}{i^2} \quad (6.27)$$

$$GLZLM_HGZE = \frac{1}{H} \sum_i \sum_j GLZLM(i, j) \cdot i^2 \quad (6.28)$$

en donde podemos observar que ahora se le da énfasis a las zonas de nivel de gris bajo y alto, respectivamente.

GLZLM_SZLGE - GLZLM_SZHGE:

El Énfasis de nivel de gris bajo de zona corta y el Énfasis de nivel de gris alto de zona corta describen la distribución de las zonas homogéneas cortas con niveles de gris bajos y altos, se calculan mediante las siguientes fórmulas

$$GLZLM_SZLGE = \frac{1}{H} \sum_i \sum_j \frac{GLZLM(i, j)}{i^2 \cdot j^2} \quad (6.29)$$

$$GLZLM_SZHGE = \frac{1}{H} \sum_i \sum_j \frac{GLZLM(i, j) \cdot i^2}{j^2} \quad (6.30)$$

en este caso y al igual que en los índices de matrices anteriores, se da énfasis a los niveles de gris bajos y altos en zonas homogéneas cortas.

GLZLM_LZLGE - GLZLM_LZHGE:

El Énfasis de nivel de gris bajo de zona larga y el Énfasis de nivel de gris alto de zona larga son el caso contrario de los dos anteriores, ya que describen la distribución de las zonas homogéneas largas con niveles de gris bajos y altos, mediante las siguientes expresiones

$$GLZLM_LZLGE = \frac{1}{H} \sum_i \sum_j \frac{GLZLM(i, j) \cdot j^2}{i^2} \quad (6.31)$$

$$GLZLM_LZHGE = \frac{1}{H} \sum_i \sum_j GLZLM(i, j) \cdot i^2 \cdot j^2 \quad (6.32)$$

y como es de esperar, dan énfasis a las zonas homogéneas largas con nivel de gris bajo y alto.

GLZLM_GLNUz - GLZLM_ZLNU:

La No-uniformidad de nivel de gris para la zona o la No-uniformidad de longitud de zona describe la falta de uniformidad de los niveles de gris y la falta de uniformidad en la longitud de las zonas homogéneas. Se calcula mediante las siguientes expresiones

$$GLZLM_GLNUz = \frac{1}{H} \sum_i \left(\sum_j GLZLM(i, j) \right)^2 \quad (6.33)$$

$$GLZLM_ZLNU = \frac{1}{H} \sum_j \left(\sum_i GLZLM(i, j) \right)^2 \quad (6.34)$$

podemos notar que se resalta la diferencia en niveles de gris y la diferencia en las longitudes de las zonas homogéneas, respectivamente.

GLZLM_ZP:

El porcentaje de zona de la GLZLM mide la homogeneidad de las zonas con nivel de gris homogéneas y se calcula como sigue

$$GLZLM_ZP = \frac{H}{\sum_i \sum_j (j \cdot GLZLM(i, j))} \quad (6.35)$$

Índices de forma

Como su nombre lo indica, describen características de forma y constan de tres índices que se definen continuación.

Esfericidad:

Describe que tan esférico es el volumen de interés, para una esfera perfecta vale 1. Se calcula mediante la siguiente ecuación

$$FORMA_Esfericidad = \frac{\pi^{1/3} \cdot (6V)^{2/3}}{A} \quad (6.36)$$

donde V y A corresponden al volumen y la superficie del Volumen de Interés basado en la triangulación de Delaunay.

Compacidad:

Describe que tan compacto es el volumen de interés y se calcula mediante la siguiente fórmula

$$FORMA_Compacidad = \frac{A^{3/2}}{V} \quad (6.37)$$

Volumen(mL y vóxeles):

Es el volumen de interés en mL o vóxeles y se expresa como sigue

$$\text{FORMA_Volumen} = \sum_i V_i \quad (6.38)$$

en donde V_i corresponde al volumen de un vóxel i que se encuentra en el volumen de interés.

Características de histograma de primer orden

Son características que describen la distribución de niveles de gris y los índices que se derivan del histograma dependen del tamaño de bin elegido, estos índices se describen a continuación.

Entropía:

Refleja la aleatoriedad de la distribución y se puede calcular aplicando la función logaritmo en base 2 y en base 10 mediante las siguientes ecuaciones

$$\text{DISCRETIZADO_HISTO_Entropía log 2} = - \sum_i p(i) \cdot \log_2(p(i) + \varepsilon) \quad (6.39)$$

$$\text{DISCRETIZADO_HISTO_Entropía log 10} = \sum_i p(i)^2 \quad (6.40)$$

en donde $p(i)$ representa la probabilidad de ocurrencia de los vóxeles con nivel de gris i y $\varepsilon = 2e^{-16}$.

Energía:

Refleja la uniformidad de la distribución y se calcula de la siguiente manera

$$\text{DISCRETIZADO_HISTO_Energía} = \sum_i p(i)^2 \quad (6.41)$$

DISCRETIZADO_AUC_CSH:

Refleja la intensidad acumulada de los histogramas ($p(i)$) producidos por un volumen parcial de una ROI, con una intensidad por encima de cierto umbral, el cual se traza contra ese valor de umbral (j), que varía de 1 al nivel de gris. Es un índice cuantitativo de heterogeneidad de captación de trazador y/o respuesta heterogénea donde menores valores corresponden a una mayor heterogeneidad. Se calcula mediante la siguiente expresión

$$\text{DISCRETIZADO_AUC_CSH} = \sum_j \sum_{i \geq j} p(i) * bw \quad (6.42)$$

en donde $p(i)$ es la probabilidad de ocurrencia de los vóxeles con nivel de gris i , j es el nivel de gris y bw es el ancho de bin.

Índices Convencionales - Discretizados

Los siguientes índices, representan características de primer orden que no están catalogadas en categorías en específico y para cada uno de estos índices, existe un valor convencional y un valor discretizado al realizar las mediciones radiómicas, sin embargo, las definiciones para

ambos casos son las mismas.

Cuartiles:

Un cuartil es un tipo de cuantil que divide el número de puntos de datos en cuatro partes aproximadamente iguales o cuartos. Debido a que los datos deben ordenarse de menor a mayor para calcular cuartiles, los cuartiles son una forma de orden estadístico. Considerando que los datos se dividen en cuatro partes iguales, se deben tener tres cuartiles a calcular:

- Q1: El primer cuartil se define como el número medio entre el número más pequeño y la mediana del conjunto de datos. Los datos que están antes de este cuartil, representan el 25 % de los datos, si estos están ordenados de menor a mayor.
- Q2: El segundo cuartil es la mediana del conjunto de datos y el 50% de los datos se encuentra por debajo de este punto.
- Q3: El tercer cuartil es el valor medio entre la mediana y el valor más alto del conjunto de datos. El 75 % de los datos se encuentra por debajo de este punto.

Intensidad Radial Media:

La intensidad radial media o RIM por sus siglas en inglés, describe la intensidad media de capas radiales con grosor de 1 vóxel, las cuales van disminuyendo de tamaño desde el exterior hasta el interior o centro de la ROI. Se pueden pensar como múltiples cascarrones con grosor de 1 vóxel al rededor de la ROI. De estas capas se pueden calcular otros parámetros estadísticos:

- RIM_{min} : Valor mínimo de intensidad de las capas.
- RIM_{stdev} : Desviación estándar de los valores de intensidad de vóxel de las capas.
- RIM_{max} : Valor máximo de intensidad de las capas.
- $RIM_{vol}(mL)$: Volumen en mililitros de todos los vóxeles de las capas.
- $RIM_{vol}(vx)$: Volumen en unidad de vóxeles de todos los vóxeles de las capas.
- RIM_{sum} : Es la suma de los valores de vóxeles de la capa.

TLG:

La glicólisis de lesión total o TLG ya se había definido de para una función continua, sin embargo al tratarse de vóxeles, debemos definir su forma discreta para entender como se calcula, a través de la siguiente expresión

$$TLG = V \cdot \frac{1}{N} \sum_i SUV_i \quad (6.43)$$

en donde V es el volumen en mL y la sumatoria corresponde al SUV_{mean} .

Oblicuidad:

Representa la asimetría de la distribución de niveles de gris y se calcula de la siguiente manera

$$\text{Oblicuidad} = \frac{\frac{1}{N} \sum_i (I(i) - \bar{I})^3}{\left(\sqrt{\frac{1}{N} \sum_i (I(i) - \bar{I})^2} \right)^3} \quad (6.44)$$

en donde $I(i)$ corresponde al número de vóxeles con intensidad i , N es el número total de vóxeles en el Volumen de Interés e \bar{I} es el promedio del nivel de gris en dicho volumen.

Curtosis:

Refleja la forma de la distribución de niveles de gris (angosta o plana) en relación con una distribución normal y se estima mediante la siguiente fórmula

$$\text{Curtosis} = \frac{\frac{1}{N} \sum_i (I(i) - \bar{I})^4}{\left(\frac{1}{N} \sum_i (I(i) - \bar{I})^2\right)^2} \quad (6.45)$$

en donde cada elemento corresponde a los mismos parámetros descritos para la oblicuidad.

Coefficiente de variación:

El coeficiente de variación es una medida de la dispersión de la distribución de niveles de gris de la ROI. Se calcula mediante la siguiente expresión

$$\text{CoefVar} = \frac{\sigma}{\mu} \quad (6.46)$$

en donde σ y μ son la desviación estándar y el promedio de la distribución, respectivamente.

Coefficiente de dispersión cuartil:

Es una alternativa al coeficiente de variación y se calcula mediante la siguiente fórmula

$$\text{CoefDispCuart} = \frac{P_{75} - P_{25}}{P_{75} + P_{25}} \quad (6.47)$$

en donde P_{25} y P_{75} son los percentiles 25 y 75 de la distribución, respectivamente.

Otros índices

- Mínimo: Representa el valor mínimo en el volumen de interés.
- Máximo: Representa el valor máximo en el volumen de interés.
- Promedio: Representa el valor promedio en el volumen de interés.
- Pico: Es el valor promedio máximo en una esfera de aproximadamente $1mL$.

Cabe mencionar que existen más parámetros radiómicos que se pueden calcular, sin embargo los anteriormente descritos son los que se estarán calculando en este trabajo para realizar el análisis de datos.