# Introducción

Hemos estado trabajando en la generación de reportes de medidas de volumetría de resonancias magnéticas de cerebro procesadas por el software Freesurfer. En la siguiente imagen se muestra el extracto de uno de ellos. En el encabezado ponemos el nombre o ID, la edad, y la región del cerebro a la cual corresponde el reporte (hemisferio izquierdo en este caso). La tabla tiene varias columnas: 1) región cerebral, 2) resultado, 3) intervalo de referencia para la edad y 4) percentil para la edad. La columna 1) contiene los nombres de las regiones cerebrales de Freesurfer dentro del hemisferio izquierdo traducidas a español. La columna 2) contiene el resultado del sujeto (sub-99999) en milímetros (mm); en este caso corresponde al grosor cortical, pero los reportes pueden ser también de área de superficie (mm2) y de volumen (mm3). La columna 3) contiene los intervalos de referencia para la edad; estos se calculan de aproximadamente 400 controles que hemos descargado de bases de datos en línea y procesado con Freesurfer (el cálculo se ajusta de acuerdo con la edad, para más detalles ver sección sobre el script ‘batches\_joinsANDlms.py’ más adelante). La columna 4) contiene el percentil para la edad de acuerdo con los intervalos de referencia (ver ‘batches\_joinsANDlms.py’ para más detalle también).

Table

Description automatically generated

# Script todo-en-uno para generar los reportes de una nueva resonancia magnética (‘recon\_report.sh’)

Hemos escrito un código o programa (script) que permite con un solo comando procesar con Freesurfer una nueva resonancia magnética de un sujeto, extraer las métricas y crear los reportes con los intervalos de referencia para la edad. El comando toma de argumentos: 1) el ID del sujeto (generalmente los nombramos de acuerdo con el sistema BIDS: sub-00000, donde los últimos dígitos se reemplazan por el número ID que sigue, por ejemplo, al día 2022/05/18 el ID que sigue es el sub-00146) y 2) la carpeta que contiene los DICOMs de la resonancia magnética.

Para ejecutar el ‘recon\_report.sh’ hay que abrir una terminal de Linux en la computadora Gateway. Una vez iniciado sesión, se puede presionar las teclas CTRL + ALT + T para abrir una terminal o alternativamente se puede dar click en la esquina inferior izquierda (icono que forma un cuadrado con puntos) para desplegar todas las aplicaciones de la computadora. Ahí se tiene que buscar la aplicación llamada ‘Terminal’ y abrirla. Se mostrará una ventana como la siguiente imagen.

Text

Description automatically generated

En esta debemos escribir:

/home/neurocogn/N1.0Dir/NezihNiegu/neuroimaging\_processor/recon\_report.sh {ID} {directorio\_DICOM}

Por ejemplo:

/home/neurocogn/N1.0Dir/NezihNiegu/neuroimaging\_processor/recon\_report.sh sub-00146 /home/neurocogn/N1.0Dir/MRI/RM\_nueva/exam/IG3QA2Q3/

Después hay que presionar ENTER para ejecutar el comando.

Noten que hay espacios después de .sh y después de 146. En BASH (el lenguaje de la terminal de linux) los espacios o la ausencia de estos son importantes. Los espacios en el comando anterior están separando el comando de sus 2 argumentos (inputs o entradas). A la izquierda del primer espacio está el comando. Este script (recon\_report.sh) se localiza en el directorio /home/neurocogn/N1.0Dir/NezihNiegu/neuroimaging\_processor/. Toma 2 argumentos, el primero está a la derecha del primer espacio y es el ID que asignamos al sujeto. Este se usará para el nombre de todas las carpetas en donde se guardarán los archivos. El segundo argumento está después del segundo espacio e indica el directorio en donde se encuentran las carpetas con los DICOMs de las diferentes secuencias de resonancia magnética del sujeto: /home/neurocogn/N1.0Dir/MRI/RM\_nueva/exam/. Generalmente guardamos las resonancia magnéticas de las que queremos obtener un reporte en una carpeta dentro de /home/neurocogn/N1.0Dir/MRI/. Para ahorrar tiempo en la búsqueda de DICOMs que hace el script, se recomienda usar como 2° argumento la carpeta que contiene muchas otras carpetas con nombres con letras en mayúsculas y números. En el caso de las resonancias magnéticas los DICOMs están en la carpeta exam/IG3QA2Q3/ donde la última carpeta tiene nombres similares y contiene varias carpetas también con nombres similares (ver imagen de abajo).

Graphical user interface, application

Description automatically generated

Hay otras ocasiones en que hay una carpeta que se llama DICOM que tiene subcarpetas con los archivos DICOMs. En este caso también se recomienda abrir las subcarpetas hasta encontrar la que tenga varias carpetas como en la imagen anterior y dar como entrada este directorio al 2° argumento del comando.

El script ‘recon\_report.sh’ realiza lo siguiente: 1) convierte los DICOMs a NIfTI, 2) procesa con Freesurfer el archivo NIfTI, 3) extrae las métricas de los archivos generados por Freesurfer, y 4) genera los reportes en HTML, PDF y CSV de cada una de las métricas y regiones. Cada uno de los pasos se guarda en una carpeta diferente y a continuación cubriremos más a detalle donde se encuentran, los archivos que estas contienen y los 4 pasos listados anteriormente.

El paso 1) busca DICOMs en todas las subcarpetas del directorio proporcionado como segundo argumento después del comando. Los convierte a formato NIfTI usando el comando dcm2niix y además extrae la edad del sujeto del header de uno de los DICOMs. Usualmente los estudios de resonancia magnética contienen varias secuencias adquiridas. El script convierte todos los DICOMs a formato NIfTI y los guarda en una carpeta nueva en /home/neurocogn/N1.0Dir/all\_niis/ con el ID del sujeto proporcionado en el 1° argumento de recon\_report.sh como nombre de la carpeta.

Para el paso 2) con Freesurfer utilizamos la secuencia T1 de los sujetos. Siempre elegimos aquella que tenga mayor tamaño de archivo y que no sea una adquisición con contraste (gadolinio) intravenoso. El script hace esto automáticamente y ejecuta el pipeline de Freesurfer con el comando recon-all con la resonancia magnética T1 elegida. Procesar una imagen T1 con Freesurfer toma aproximadamente 10 horas. Freesurfer guarda todos los archivos que genera predeterminadamente en el directorio /home/neurocogn/N1.0Dir/freesurfer/subjects/ dentro de una carpeta con el ID alimentado como 1° argumento del script recon\_report.sh

En el paso 3), de los archivos resultantes del procesamiento extraemos las métricas de volumen (mm3) de las regiones subcorticales del cerebro, del hipocampo, del tallo cerebral y de los hemisferios cerebrales izquierdo y derecho; de estos 2 últimos también se puede extraer el grosor cortical (mm) y el área de superficie (mm2). Estos se extraen en los archivos .txt que se guardan en /home/neurocogn/N1.0Dir/NezihNiegu/neuroimaging\_processor/Volumetry/ dentro de una carpeta nueva con nombre del ID del sujeto proporcionado en el 1° argumento. Estos archivos contienen las diferentes estructuras de las regiones del cerebro en columnas y su respectiva métrica; todos los valores separados por tabulaciones o un espacio en blanco.

Finalmente, en el paso 4), estos archivos .txt se utilizan para generar los reportes por medio de un script de Python (createPDFreport\_spanish.py) que ejecuta recon\_report.sh. Este script de Python, genera los reportes en HTML y PDF de cada una de las métricas y regiones (ver imagen en sección de introducción) y los guarda en la misma carpeta de los archivos .txt. Para las últimas columnas del reporte, extrae los intervalos de referencia para la edad de los archivos encontrados en la carpeta /home/neurocogn/N1.0Dir/NezihNiegu/neuroimaging\_processor/Volumetry/batches/prediction\_intervals/. Estos corresponden a archivos pickle (.pkl) que contienen diccionarios de Python con los intervalos ya guardados para cada edad. Los archivos .pkl se generaron con otro script de Python (batches\_joinsANDlms.py) que conjunta todas las métricas de las resonancias magnéticas de controles que descargamos de bases de datos en línea y procesamos con Freesurfer. Utiliza una regresión lineal (y sus intervalos de predicción al 95%) de cada métrica de cada estructura cerebral en función de la edad. Estos 2 scripts de Python se encuentran en /home/neurocogn/N1.0Dir/NezihNiegu/neuroimaging\_processor/Volumetry/python\_scripts/ y describiremos más adelante cómo funcionan a mayor detalle.

Los archivos HTML y PDF están nombrados de acuerdo con la región del cerebro y la métrica que reportan. También hay unos CSV por si llegan a ocupar en este formato los resultados del sujeto y su percentil de acuerdo con la edad. Estos se pueden fusionar con un script de Python similar a csv\_fusion.ipynb que se encuentra en la carpeta /home/neurocogn/N1.0Dir/NezihNiegu/neuroimaging\_processor/Volumetry/jupyterScripts/. Todos los archivos están nombrados de la siguiente manera: \*\_región\_métrica.pdf, \*\_región\_métrica.html, \*\_región\_métrica.csv donde región puede ser “brainstem=tallo cerebral”, “hippo=hipocampo”, “lh=hemisferio izquierdo”, “rh=hemisferio derecho”, “seg=estructuras subcorticales” y donde métrica puede ser “vol=volumen”, “area=área de superficie”, “thick=grosor cortical”.

# Detalles del script de Python ‘createPDFreport\_spanish.py’

Para usar este script de Python se debe tener una carpeta con los archivos .txt del paso 3) descrito en la sección anterior de recon\_report.sh. El script ‘createPDFreport\_spanish.py’ tiene documentación que se puede consultar desde la terminal de Python. Para abrir Python y poder ejecutar el script o consultar su documentación sólo hay que escribir y ejecutar los siguientes 2 comandos en la terminal de Linux (ver sección anterior para ver como abrir una terminal de Linux en la computadora Gateway):

cd /home/neurocogn/N1.0Dir/NezihNiegu/neuroimaging\_processor/Volumetry/python\_scripts/

python3

Hay que dar ENTER después de escribir el primer comando para después escribir el segundo comando y dar ENTER de nuevo. Esto abrirá una sesión de Python 3 en la terminal con el directorio de trabajo en establecido en la carpeta del 1° comando.

Lo primero que hay que hacer en la sesión de Python es importar el script ‘createPDFreport\_spanish.py’. Para esto hay que ejecutar el siguiente comando (escribir y luego presionar ENTER):

from createPDFreport\_spanish import \*

Después de esto se podrá consultar la documentación al ejecutar el siguiente comando:

help(createPDFreport\_spanish)

Esto mostrará la documentación en inglés del script. Para salir de esta hay que presionar la tecla ‘q’. En resumen, el script toma 5 argumentos:

1. patient\_age: corresponde a la edad del sujeto (se escribe el número entero)
2. region: región de Freesurfer de la cual se quiere generar el reporte (se escribe cualquiera de las siguientes opciones: "lh" para hemisferio izquierdo, "rh" para hemisferio derecho, "seg" para regiones subcorticales, "brainstem" para el tallo cerebral, "hippo" para el hipocampo)
3. subjectANDoutput\_dir: es el directorio de la carpeta donde se encuentran los archivos .txt que se generan con el paso 3) del script recon\_report.sh y también donde se crearan los archivos HTML, PDF y CSV con la información del reporte de la sección de introducción
4. metric: métrica de la cual se quiere generar el reporte (se escribe cualquiera de las siguientes opciones: "vol" para volumen, "area" para área de superficie, "thick" para grosor cortical; la opción predeterminada es “vol”)
5. prediction\_intervals\_dir: es el directorio de la carpeta donde se encuentran los intervalos de referencia en archivos .pkl de los aproximadamente 400 sujetos controles procesados y que generamos con el script de Python ‘batches\_joinsANDlms.py’ (el funcionamiento de este se describirá más a detalle en la siguiente sección; el directorio predeterminado es '/home/neurocogn/N1.0Dir/NezihNiegu/neuroimaging\_processor/Volumetry/batches/prediction\_intervals/')

Por ejemplo, para generar los reportes con el volumen de las estructuras subcorticales de un sujeto de 61 años (sub-00024\_2020) se ejecutaría este comando en la sesión de Python:

createPDFreport\_spanish(patient\_age=61, region="seg", subjectANDoutput\_dir="/home/neurocogn/N1.0Dir/NezihNiegu/neuroimaging\_processor/Volumetry/sub-00024\_2020/", metric='vol', prediction\_intervals\_dir='/home/neurocogn/N1.0Dir/NezihNiegu/neuroimaging\_processor/Volumetry/batches/prediction\_intervals/')

El script ‘createPDFreport\_spanish.py’ básicamente extrae los intervalos de referencia de los diccionarios de Python guardados en los archivos .pkl y los resultados del sujeto de los .txt y los acomoda en una tabla. Para los nombres en español, creamos archivos CSV con los nombres de las regiones de Freesurfer escritas en español en: /home/neurocogn/N1.0Dir/NezihNiegu/neuroimaging\_processor/Volumetry/python\_scripts/CSVs\_with\_report\_names/ y de estos CSV se extraen los nombres que se colocan en la tabla. También, con los intervalos de referencia se calcula el percentil al que el resultado del sujeto corresponde para reportarlo en la tabla. Se genera un HTML con la tabla, el cual se convierte a PDF. El script también genera un archivo CSV reportando el resultado del sujeto y su percentil. Ver último párrafo en la sección anterior para entender los nombres de los archivos que se crean.

# Detalles del script de Python ‘batches\_joinsANDlms.py’

Al igual que el script anterior, ‘batches\_joinsANDlms.py’ tiene documentación que se puede consultar desde la terminal de Python. Para abrir Python y poder ejecutar el script o consultar su documentación sólo hay que ejecutar los siguientes 2 comandos en la terminal de Linux (ver sección del script ‘recon\_report.sh’ para ver cómo abrir una terminal de Linux en la computadora Gateway):

cd /home/neurocogn/N1.0Dir/NezihNiegu/neuroimaging\_processor/Volumetry/python\_scripts/

python3

Para importar el script a la sesión de Python hay que ejecutar el siguiente comando:

from batches\_joinsANDlms import \*

Después de esto se podrá consultar la documentación al ejecutar el siguiente comando:

help(batches\_joinsANDlms)

Esto mostrará la documentación en inglés del script. Para salir de esta hay que presionar la tecla ‘q’. En resumen, el script toma 4 argumentos:

1. group\_age: grupo de edad para los intervalos de predicción; dividimos los aproximadamente 400 controles en 2 grupos, uno mayor o igual a 40 años y otro menor de 40 años (se escribe "40ormoreyears" para el grupo ≥40 años o "lessthan40years" para el grupo <40 años)
2. region: región de Freesurfer de la cual se quiere generar los intervalos de predicción (se escribe cualquiera de las siguientes opciones: "lh" para hemisferio izquierdo, "rh" para hemisferio derecho, "seg" para regiones subcorticales, "brainstem" para el tallo cerebral, "hippo" para el hipocampo)
3. metric: métrica de la cual se quiere generar el reporte (se escribe cualquiera de las siguientes opciones: "vol" para volumen, "area" para área de superficie, "thick" para grosor cortical; la opción predeterminada es “vol”)
4. batches\_dir: directorio que contiene los .txt de todos los aproximadamente 400 sujetos controles y que contiene una carpeta llamada ‘prediction\_intervals’ donde se guardará el archivo pickle (.pkl) con el diccionario de Python que contiene los intervalos de predicción de acuerdo con la edad (opción predeterminada es "/home/neurocogn/N1.0Dir/NezihNiegu/neuroimaging\_processor/Volumetry/batches/")

Por ejemplo:

batches\_joinsANDlms(group\_age="lessthan40years", region="brainstem", metric="vol")

Este ejemplo anterior lo que hace es generar los archivos .pkl que contienen un diccionario de Python con los intervalos de predicción de volumen (“vol”) de aquellos sujetos controles que tenemos menores a 40 años (“lessthan40years”). Estos para cada una de las subregiones en el tallo cerebral (“brainstem”). Como no alimentamos a la función una opción a “batches\_dir”, toma la opción predeterminada (ver punto 4).

En el directorio "/home/neurocogn/N1.0Dir/NezihNiegu/neuroimaging\_processor/Volumetry/batches/" de la computadora Gateway hay un CSV con la edad de todos los controles que procesamos con Freesurfer. Las tandas (batches) de sujetos procesados están guardados en diferentes archivos .txt. La base de datos ICBM fue la primera que procesamos y estos .txt están guardados directamente en la carpeta anterior. Los batches siguientes los fuimos procesando en tandas de 60 sujetos. Están ubicados en 4 carpetas: batch\_1, batch\_2, batch\_3, batch\_4; cada carpeta contiene 2 subcarpetas: folder\_1, folder\_2. El script selecciona todos los archivos que corresponden a la región y a la métrica proporcionadas en los argumentos (‘region’ y ‘metric’). Aquellos .txt que contienen la información de hipocampo (“hippo”) y tallo cerebral (“brainstem”) se encuentran directamente en la carpeta mencionada al principio de párrafo. De la misma manera, el batch correspondiente a la base de datos ICBM se encuentra directamente en la carpeta mencionada al principio bajo los nombres \*\_ICBM\_\*.txt .

Todos estos .txt se cargan a Python como dataframes de pandas y se fusionan por medio de joins entre sí y con la edad y género cuya información se encuentra en el CSV mencionado en el párrafo anterior. Finalmente se hace una regresión lineal en función de la edad de cada subregión de la región y la métrica seleccionada. Los intervalos de predicción al 95% de la regresión lineal son calculados para cada una de las edades dentro del grupo de edad seleccionado: de 15 a 39 para "lessthan40years" y de 40 a 100 para "40ormoreyears". El límite inferior y superior son guardados en un diccionario que contiene la edad en sus llaves (keys del diccionario). Dentro de cada key está un array con el nombre de la subregión, el límite inferior y el límite superior de su intervalo de predicción para esa edad indicada en el key. Este diccionario se guarda en la carpeta ‘prediction\_intervals’ como un archivo pickle (.pkl) con el nombre indicando la región, métrica y grupo de edad seleccionados. Al momento ya se generaron todos los archivos .pkl de todos los sujetos controles que se han procesado hasta ahora. Si se procesan otros batches habrá que modificar el script para que este incluya estos nuevos datos y después de eso generar los nuevos archivos .pkl que contengan los nuevos intervalos de predicción.

# Directorios en computadora Gateway con sujetos controles procesados

La mayoría de los sujetos controles fueron procesados en esta computadora antes de que la formateáramos. Cuando se procesa una resonancia magnética se crean todos los archivos generados por Freesurfer en el directorio contenido en la variable $SUBJECTS\_DIR . Al formatear la computadora reinstalamos Ubuntu 18 en otra partición y el antiguo $SUBJECTS\_DIR quedó en el siguiente directorio:

/media/neurocogn/ee9bf889-122a-432e-b4bc-6235e61c7577/usr/local/freesurfer/subjects

En este directorio se encuentran la mayoría de los archivos de Freesurfer de la mayoría de los sujetos controles que hemos procesado. El 1er batch se encuentra en la carpeta ‘dir\_with\_Treconall\_1stbatch’ y proviene del procesamiento de las resonancias magnéticas que están en la carpeta ‘dir\_with\_T1MRI\_1stbatch’. El 2do y 3er batch también están en carpetas similares: ‘dir\_with\_Treconall\_\*batch’ y ‘dir\_with\_T1MRI\_\*batch’. El 4to batch está directamente en el directorio del párrafo anterior. Corresponde a las carpetas desde ‘ADNI\_022…’ hasta ‘ADNI\_141…’ y estas carpetas contienen los archivos generados por Freesurfer del procesamiento de las resonancias magnéticas en la carpeta ‘dir\_with\_T1MRI’. Todos estos batches corresponden a sujetos controles de las bases de datos PPMI y ADNI.

También tenemos resonancias magnéticas de la base de datos ICBM que se encuentran en el $SUBJECTS\_DIR de Ubuntu 18 después de que formateamos la computadora. Este corresponde al directorio:

/home/neurocogn/N1.0Dir/freesurfer/subjects

Aquí están las carpetas desde ‘ICBM\_MNI\_0103…’ hasta ‘ICBM\_MNI\_0349…’ y provienen de los DICOMs que están en las carpetas localizadas en estos directorios: /home/neurocogn/N1.0Dir/NezihNiegu/neuroimaging\_processor/Volumetry/batches/ICBM/ICBM\_\*/ICBM/

Hemos procesado las bases de datos PPMI, ICBM y empezado con ADNI. En total llevamos aproximadamente 400 sujetos. Seleccionamos las resonancias magnéticas T1 tridimensionales que fueran de escáneres de 1.5 o 3 Teslas. De ADNI solo hemos procesado aquellos sujetos dentro de la carpeta del directorio:

/media/neurocogn/ee9bf889-122a-432e-b4bc-6235e61c7577/home/neurocogn/Documents/Victor\_Issa-Garcia/ADNI/ADNI\_T1\_controls\_1.5T

Estos se encuentran en el ‘folder2’ del batch3 de $SUBJECTS\_DIR antes de formatear la computadora. También corresponden al batch4 (ver 2° párrafo de esta sección). Dentro de la carpeta del siguiente directorio se encuentran los controles de ADNI 1.5 Tesla faltantes:

$SUBJECTS\_DIR/dir\_with\_T1MRI\_3rdbatch/ADNI\_pendientes

En caso de decidir continuar procesando más sujetos, después de procesar esos 6 sujetos. Habría que continuar con los restantes de ADNI 3 Tesla que hemos descargado y se encuentran en las carpetas ‘ADNI’, ‘ADNI (2)’, ‘ADNI (3)’, ‘ADNI (4)’ y ‘ADNI (5)’ dentro de:

/media/neurocogn/ee9bf889-122a-432e-b4bc-6235e61c7577/home/neurocogn/Documents/Victor\_Issa-Garcia/ADNI/

La estructura de carpetas dentro de estas 5 carpetas es la que se descarga directamente del portal de IDA LONI (<https://ida.loni.usc.edu/login.jsp>). Reestructuramos el nombre de las subcarpetas y archivos usando el script ‘ADNI\_cp-into-proper-struc.sh’ en la carpeta del directorio:

/media/neurocogn/ee9bf889-122a-432e-b4bc-6235e61c7577/usr/local/freesurfer/subjects/scripts

Para continuar con ADNI 3 Tesla habrá que modificar este script para reestructurar el nombre de los archivos y las carpetas. En esta carpeta ‘scripts’ también tenemos los scripts ‘PPMI\_cp-into-proper-struc.sh’ y ‘PPMI\_3T\_cp-into-proper-struc.sh’ que fueron usados para copiar los DICOMs de la base de datos PPMI a la estructura de las carpetas ‘dir\_with\_T1MRI\_1stbatch’, ‘dir\_with\_T1MRI\_2ndbatch’ y ‘dir\_with\_T1MRI\_3rdbatch’ en $SUBJECTS\_DIR antes de formatear la computadora.

# Procesamiento de sujetos controles en los directorios anteriores (folder\_recon.sh)

En el siguiente directorio se encuentra la carpeta ‘scripts’ se encuentra un script con el nombre de ‘folder\_recon.sh’:

/media/neurocogn/ee9bf889-122a-432e-b4bc-6235e61c7577/usr/local/freesurfer/subjects

En el script, hay que cambiar el directorio asignado a la variable ‘folder\_to\_recon’. Se debe asignar el directorio de una carpeta que contenga carpetas con el nombre de cada sujeto a procesar con Freesurfer. Dentro de cada carpeta del sujeto deben encontrarse los archivos DICOMs de la resonancia magnética T1. El script entonces procesará todos los sujetos dentro del directorio asignado a ‘folder\_to\_recon’. Los archivos de Freesurfer se guardan en $SUBJECTS\_DIR en carpetas con el mismo nombre de las carpetas de los sujetos en ‘folder\_to\_recon’. Al finalizar de procesar todos los sujetos con Freesurfer se creará en ‘folder\_to\_recon’ una carpeta nueva en con nombre ‘stats\_\*’ que contiene los archivos .txt con las métricas de todos los sujetos procesados (lo que llamamos tandas o batches en la sección del script ‘batches\_joinsANDlms.py’).