

به نام خدا



دانشگاه شهید بهشتی
دانشکده مهندسی و علوم کامپیوتر
رشته مهندسی کامپیوتر

معرفی رویکردی یکپارچه برای ذخیره‌سازی و پایش داده‌های مغزی در بستر BIDS

گزارش پروژه کارشناسی

دانشجو:
مبینا شهبازی

استاد راهنما:
دکتر حسن حقیقی

تابستان ۱۴۰۴

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

سپاسگزاری

سپاس و قدردانی خود را نخست تقدیم می‌کنم به استاد ارجمندم، جناب آقای دکتر حسن حقیقی، که راهنمایی‌های ارزشمند و حمایت‌های علمی ایشان نقش تعیین‌کننده‌ای در پیشبرد این پژوهش داشت. همچنین این پایان‌نامه را با افتخار به پدر و مادر مهربانم تقدیم می‌کنم؛ بزرگوارانی که با صبوری و پشتیبانی همیشگی خود مسیر تلاش و پیشرفتم را هموار ساخته‌اند.

چکیده

در سال‌های اخیر، رشد چشمگیر داده‌های چندوجهی در پروژه‌های بزرگ تصویربرداری مغزی، نیاز به روش‌های کارآمد برای مدیریت، ذخیره‌سازی و جستجو را بیش از پیش آشکار کرده است. استاندارد BIDS با هدف سازمان‌دهی و یکپارچه‌سازی داده‌های تصویربرداری معرفی شد، اما در مقیاس بایوبانک‌ها، همچنان چالش‌هایی مانند سرعت جستجو، مصورسازی تعاملی و به‌روزرسانی پویا پابرجاست.

این پروژه با هدف ارائه یک اثبات مفهومی، چارچوبی یکپارچه برای مدیریت داده‌های BIDS طراحی و پیاده‌سازی کرده است. در این چارچوب، داده‌ها و فراداده‌ها با استفاده از کتابخانه PyBIDS استخراج و پس از ایندکس‌گذاری در Elasticsearch، از طریق Kibana در قالب داشبوردهای تحلیلی مصورسازی شدند. برای تضمین پویایی سامانه و انطباق با شرایط واقعی بایوبانک‌ها، فرایند ایندکس‌گذاری و به‌روزرسانی با Apache Airflow خودکارسازی گردید و کلیه سرویس‌ها در بستر Docker یکپارچه‌سازی شدند.

ارزیابی سامانه بر اساس سه سناریو طراحی شد: (۱) افزودن داده جدید و بررسی پویایی در انعکاس تغییرات، (۲) مقایسه عملکرد PyBIDS و Elasticsearch در جستجوی داده‌ها از نظر سرعت و صحت، و (۳) آزمون پایداری در اجرای خودکار گردش کارها. نتایج نشان داد که سامانه توانایی مدیریت افزوده‌شدن داده‌های جدید، سرعت بازیابی حدود ۸۳۰ برابر سریع‌تر نسبت به روش پایه، و پایداری قابل قبول در به‌روزرسانی‌های دوره‌ای دارد.

در مجموع، روش پیشنهادی توانست نیازمندی‌های اصلی پروژه را برآورده سازد و به‌عنوان یک راهکار مقیاس‌پذیر و عملی برای مدیریت داده‌های پیچیده تصویربرداری مغزی مطرح گردد.

واژگان کلیدی: داده‌های چندوجهی مغزی، استاندارد ساختار داده‌های تصویربرداری مغزی، ایندکس‌گذاری، الاستیک‌سرچ^۱، مصورسازی داده‌ها

^۱ Elasticsearch

فهرست مطالب

فصل اول: کلیات	۱۰
۱-۱ مقدمه	۱۱
۲-۱ بیان مسئله	۱۲
۳-۱ کلیات روش پیشنهادی	۱۳
۴-۱ ساختار گزارش	۱۴
فصل دوم: مفاهیم پایه و کارهای مرتبط	۱۵
۱-۲ مقدمه	۱۶
۲-۲ مفاهیم پایه	۱۶
۱-۲-۲ داده‌های تصویربرداری مغز و انواع آن	۱۶
۲-۲-۲ فرمت‌ها و استانداردهای ذخیره‌سازی داده‌ها	۱۸
۳-۲-۲ ابزارها و کتابخانه‌های مرتبط با BIDS	۲۲
۴-۲-۲ ایندکس‌گذاری و جست‌وجو در داده‌ها	۲۴
۳-۲ تحلیل نقاط قوت و ضعف کارهای مشابه	۲۵
۴-۲ جمع‌بندی	۲۷
فصل سوم: روش پیشنهادی و نتیجه‌گیری	۲۸
۱-۳ مقدمه	۲۹
۲-۳ ساختار روش پیشنهادی	۲۹
۱-۲-۳ نیازمندی‌های سیستم	۳۰
۲-۲-۳ نمای کلی سیستم و معماری پیشنهادی	۳۱
۳-۲-۳ فازبندی اجرای پروژه	۳۴
۴-۲-۳ جریان داده‌ها در سیستم	۳۵
۳-۳ پیاده‌سازی روش پیشنهادی	۳۶
۱-۳-۳ استخراج و پردازش داده و فراداده	۳۶
۲-۳-۳ ایندکس‌گذاری و ذخیره‌سازی	۳۸
۳-۳-۳ پایش و مصورسازی داده‌ها	۳۹

- ۴۱..... ۴-۳-۳ رابط کاربری برای افزودن داده‌های جدید
- ۴۲..... ۵-۳-۳ ماژول به‌روزرسانی خودکار با Airflow
- ۴۳..... ۶-۳-۳ یکپارچه‌سازی ماژول‌ها در زیرساخت Docker
- ۴۴..... ۴-۳ روش ارزیابی
- ۴۴..... ۱-۴-۳ نوع داده‌ها و مجموعه داده مورد استفاده برای ارزیابی
- ۴۵..... ۲-۴-۳ سناریوهای ارزیابی عملکرد
- ۵۵..... ۵-۳ نتیجه‌گیری
- ۵۷..... ۶-۳ جمع‌بندی

فهرست شکل‌ها

- شکل ۱- نمای کلی فرایند سیستم ۱۴
- شکل ۲- نمونه‌ای ساده از یک ساختار گزارش DICOM ۱۸
- شکل ۳- ساختار یک مجموعه داده سازگار با BIDS ۲۰
- شکل ۴- خروجی‌های آزمایش‌های تصویربرداری عصبی (چپ) و تبدیل آن به مجموعه داده سازگار با BIDS ۲۱
- شکل ۵- فرایند تبدیل داده خام به قالب BIDS ۲۲
- شکل ۶- استخراج فراداده با کتابخانه PyBIDS ۲۳
- شکل ۷- تشریح نحوه پشتیبانی از مدالیته EEG در نسخه جدید BIDS ۲۵
- شکل ۸- نمودار موارد کاربری ۳۱
- شکل ۹- نمای معماری کلان سیستم ۳۳
- شکل ۱۰- جریان دادگان از تصویربرداری تا مصورسازی ۳۶
- شکل ۱۱- بخشی از فرایند استخراج فراداده فایل‌ها ۳۷
- شکل ۱۲- انتقال دادگان استخراج‌شده به Elasticsearch ۳۸
- شکل ۱۳- نمای داشبورد نخست (اطلاعات دموگرافیک آزمودنی‌ها) ۳۹
- شکل ۱۴- نمای داشبورد دوم (فراداده فایل‌ها) ۴۰
- شکل ۱۵- نمای داشبورد سوم (محتوای فایل‌ها) ۴۰
- شکل ۱۶- رابط کاربری برای افزودن داده ۴۱
- شکل ۱۷- رابط کاربری سرویس Apache Airflow ۴۲
- شکل ۱۸- نمونه‌ای از تعریف یک گردش کار ۴۳
- شکل ۱۹- مجموعه‌داده خام پیش از افزودن داده ۴۶
- شکل ۲۰- پرس‌وجو برای شمارش آزمودنی‌ها در در ایندکس subject_6040 ۴۷
- شکل ۲۱- نمای داشبورد آزمودنی‌ها ۴۷
- شکل ۲۲- افزودن داده از طریق رابط کاربری ۴۸
- شکل ۲۳- مجموعه‌داده خام پس از افزودن داده ۴۹
- شکل ۲۴- اجرای گردش کار آزمودنی‌ها در Airflow ۵۰
- شکل ۲۵- پرس‌وجو برای شمارش آزمودنی‌ها در در ایندکس subject_6040 ۵۰

شکل ۲۶ - نمای داشبورد آزمودنی‌ها	۵۱
شکل ۲۷ - کد اجرای پرس‌وجو نمونه با استفاده از کتابخانه PyBIDS	۵۲
شکل ۲۸ - نتیجه اجرای پرس‌وجو نمونه با استفاده از کتابخانه PyBIDS	۵۲
شکل ۲۹ - پرس‌وجو برای شمارش آزمودنی‌ها پس از افزودن داده	۵۳
شکل ۳۰ - روند اجرای دوره‌ای	۵۴

فهرست جداول

جدول ۱- مقایسه نتایج دو روش جستجو با PyBIDS و Elasticsearch	۵۳
جدول ۲- نتایج ارزیابی	۵۶

فهرست کلمات اختصاری

Abbreviations	Pages numbers
BIDS: Brain Imaging Data Structure	2
HCP: Human Connectome Project	2
EEG: Electroencephalography	2,3,8
fMRI: Functional Magnetic Resonance Imaging	2,3,8
MEG: Magnetoencephalography	2,3,8,10
sMRI: Structural Magnetic Resonance Imaging	7
NIfTI: Neuroimaging Informatics Technology Initiative	7,10
CT: Computed Tomography	8
NIRS: Near-Infrared Spectroscopy	8
DTI: Diffusion Tensor Imaging	10
PET: Positron Emission Tomography	16
DAG: Directed Acyclic Graph	23
DWI: Diffusion-Weighted Imaging	28,35

فصل اول: کلیّات

۱-۱ مقدمه

در دهه‌های اخیر، پیشرفت فناوری‌های تصویربرداری عصبی، امکان تولید حجم عظیمی از داده‌های مغزی را فراهم کرده است. داده‌هایی که از مدالیت‌های مختلف مانند MRI، fMRI، EEG، MEG و ... به دست می‌آیند، در پروژه‌های تحقیقاتی، کلینیکی و بایوبانک‌ها نقشی کلیدی دارند. در برخی پروژه‌های بزرگ تصویربرداری مانند پروژه اتصال مغزی انسان^۲ در ابتدای پروژه تخمین زده می‌شد حجم داده‌های پردازش‌شده به حدود یک پتابایت برسد [۱]، اما تا به امروز، بیش از ۲۷ پتابایت داده به اشتراک گذاشته شده است [۲] همچنین، بانک اطلاعاتی عظیمی مانند بانک زیستی بریتانیا^۳ در مجموع بیش از ۳۰ پتابایت داده شامل داده‌های تصویربرداری، ژنتیکی و سلامت دارد و این حجم به‌طور مداوم در حال افزایش است [۳].

با وجود ارزش بالای این داده‌ها، مدیریت، ذخیره‌سازی، و بازیابی آن‌ها همواره یک چالش بزرگ بوده است. تا سال‌ها، قالب DICOM به‌عنوان استاندارد اصلی ذخیره‌سازی داده‌های تصویربرداری پزشکی استفاده می‌شد؛ اما برای تحلیل‌های علمی در حوزه علوم اعصاب، این قالب به دلیل پراکندگی اطلاعات و ساختار غیریکپارچه، مشکلاتی ایجاد می‌کرد. برای رفع این مشکلات، استاندارد ساختار داده‌های تصویربرداری مغز^۴ معرفی شد که با سازمان‌دهی و نام‌گذاری یکپارچه فایل‌ها، دسترسی و تحلیل داده‌های مغز را تسهیل می‌کند [۴].

با افزایش ابعاد و پیچیدگی داده‌های چندوجهی در بایوبانک‌ها، نیاز به روش‌های بهینه برای جستجو و بازیابی سریع داده‌ها بیش از پیش احساس می‌شود. به‌عنوان مثال، در یک بایوبانک ممکن است هزاران بیمار با چندین نوع داده تصویربرداری و فراداده^۵ مرتبط ثبت شده باشند که بدون یک سیستم جستجوی کارآمد، استخراج اطلاعات هدفمند از میان آن‌ها بسیار زمان‌بر خواهد بود.

پروژه حاضر با هدف بررسی ساختار BIDS و استفاده از ابزارهای نوین مدیریت داده، شامل PyBIDS برای استخراج فراداده، Elasticsearch برای ایندکس‌گذاری و جستجوی سریع، و Kibana برای ساخت داشبوردهای تحلیلی، طراحی شده است. رویکرد کلی این پروژه، ایجاد یک نمونه اولیه از ترکیب این ابزارها برای مدیریت داده‌های مغزی است که قابلیت به‌روزرسانی پویا و نمایش گرافیکی اطلاعات کلیدی بایوبانک را داشته باشد.

^۲ Human Connectome Project (HCP)

^۳ UK Biobank

^۴ Brain Imaging Data Structure (BIDS)

^۵ Metadata

۱-۲ بیان مسئله

با رشد چشمگیر پروژه‌های بزرگ تصویربرداری مغز، نیاز به مدیریت، ذخیره‌سازی و جستجوی کارآمد داده‌های حجیم چندوجهی^۶ بیش از پیش احساس می‌شود. این داده‌ها شامل مدالیت‌های گوناگونی مانند fMRI، MRI، ساختاری، MEG و EEG هستند که هر کدام فرمت‌ها، حجم‌ها و فراداده‌های مخصوص به خود را دارند.

با وجود استانداردسازی ساختار این داده‌ها در قالب BIDS، همچنان چند چالش اساسی پابرجاست:

۱. جستجو و بازیابی سریع داده‌ها با توجه به حجم عظیم و پیچیدگی ساختار.
 ۲. ابزارهای تحلیلی و داشبوردهای یکپارچه که بتوانند فراداده‌های حیاتی مانند توزیع جمعیت‌شناختی، نوع مدالیت‌ها یا حجم داده را به‌صورت تعاملی نمایش دهند.
 ۳. دشواری نگهداری سیستم‌ها در شرایط دینامیک، که در آن داده‌های جدید به‌طور مداوم به پایگاه داده افزوده می‌شوند.
- در حال حاضر، ابزارهایی مانند PyBIDS برای پیمایش داده‌ها و Elasticsearch و Kibana برای جستجو و تحلیل وجود دارند، اما اغلب استفاده‌ها از این ابزارها به صورت مجزا و غیر یکپارچه انجام می‌شود و کمتر پروژه‌ای این سه را به‌طور کامل در یک سیستم عملیاتی برای مدیریت داده‌های BIDS به کار گرفته است.
- پروژه حاضر با هدف ارائه یک اثبات مفهومی^۷، رویکردی ترکیبی برای استخراج فراداده، ایندکس‌گذاری کارآمد در Elasticsearch و ساخت داشبوردهای گرافیکی در Kibana پیشنهاد می‌دهد. این رویکرد می‌تواند به بایوبانک‌های داده مغزی کمک کند تا داده‌های خود را با کارایی بالاتر مدیریت و تحلیل کنند.

^۶ Multi-modal

^۷ Proof of Concept

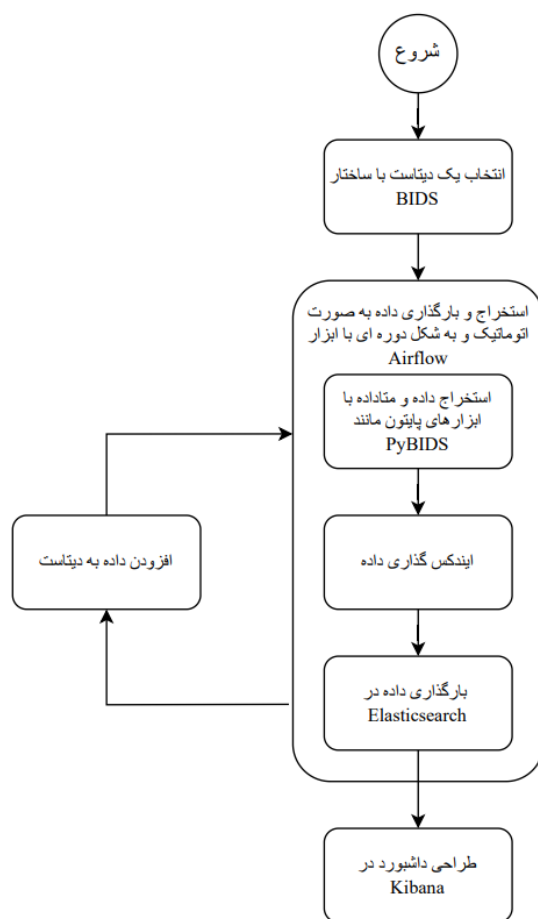
۱-۳ کلیات روش پیشنهادی

در این پروژه هدف اصلی، طراحی و پیاده‌سازی یک چارچوب یکپارچه برای مدیریت، جستجو و تحلیل فراداده‌های موجود در مجموعه داده‌های ساختار BIDS^۸ است؛ به گونه‌ای که بتواند در یک محیط پویا مانند بایوبانک داده‌های مغزی، به صورت مستمر به‌روزرسانی و استفاده شود. در گام نخست، یک مجموعه داده نمونه از منابع عمومی مانند OpenNeuro انتخاب می‌شود [۵] سپس با استفاده از کتابخانه PyBIDS و سایر ابزارهای پایتون، اطلاعات مورد نیاز مانند شناسه آزمودنی^۹، مدالیت‌ها، حجم فایل‌ها و محتوای داخل فایل‌ها استخراج می‌شود. این داده‌ها سپس به فرمت JSON سازگار با Elasticsearch تبدیل شده و با تعریف یک نگاشت ایندکس^{۱۰} مناسب که انواع داده‌ها (عددی، متنی و زمانی) را در نظر می‌گیرد، در Elasticsearch ایندکس‌گذاری می‌شوند. پس از این مرحله، قابلیت جستجو و فیلترگذاری سریع روی داده‌ها فراهم شده و بستر لازم برای تحلیل و پایش اطلاعات ایجاد می‌شود. در ادامه، با بهره‌گیری از Kibana، داشبوردهای گرافیکی متنوعی برای نمایش شاخص‌هایی مانند تعداد آزمودنی‌ها، توزیع سنی، فراوانی مدالیت‌ها و حجم داده‌ها طراحی می‌گردد. در نهایت، به منظور داینامیک‌سازی سیستم و انطباق با شرایط واقعی بایوبانک، که در آن داده‌های جدید به طور روزانه افزوده می‌شوند، از Apache Airflow برای زمان‌بندی و اجرای خودکار فرآیند استخراج و ایندکس‌گذاری استفاده می‌شود و تمام سرویس‌ها در بستر Docker استقرار می‌یابند. این روش پیشنهادی یک نقشه راه کامل از دریافت داده خام تا ارائه داشبورد تحلیلی و به‌روزرسانی خودکار را فراهم می‌آورد و می‌تواند به عنوان یک اثبات مفهومی برای پیاده‌سازی بایوبانک‌های مقیاس بزرگ مورد استفاده قرار گیرد. نمای کلی روش پیشنهادی در شکل ۱ قابل مشاهده است.

^۸ Dataset

^۹ Subject

^{۱۰} Index Mapping



شکل ۱- نمای کلی فرایند سیستم

۴-۱ ساختار گزارش

پس از آشنایی با مقدمات موضوع و درک دقیق تر مسئله در فصل نخست، در فصل دوم به تشریح مفاهیم پایه مرتبط با پروژه پرداخته خواهد شد. همچنین، پژوهش‌های مشابه مورد بررسی و تحلیل قرار می‌گیرند تا نقاط قوت و ضعف آن‌ها مشخص شود. در ادامه، فصل سوم به شرح گام‌به‌گام روش پیشنهادی اختصاص دارد؛ به این ترتیب که ابتدا ساختار کلی راهکار معرفی شده و سپس مراحل پیاده‌سازی آن توضیح داده می‌شود. در پایان این فصل نیز، روش ارزیابی سامانه و نتایج به‌دست‌آمده ارائه خواهد شد.

فصل دوم: مفاهیم پایه و کارهای مرتبط

۱-۲ مقدمه

فصل دوم این پژوهش با هدف ایجاد یک درک مشترک و دقیق از مفاهیم و فناوری‌های مورد استفاده در پروژه آغاز می‌شود. در بخش نخست، مفاهیم پایه‌ای که درک آن‌ها پیش‌نیاز فهم سازوکارها و مراحل عملیاتی پروژه است، معرفی و تشریح شده‌اند؛ این مفاهیم شامل ماهیت داده‌های تصویربرداری مغزی و انواع آن، فرمت‌ها و استانداردهای ذخیره‌سازی (DICOM، NIfTI و BIDS)، ابزارها و کتابخانه‌های پشتیبان پردازش و سازماندهی داده‌ها (dcm2niix، HeuDiConv، PyBIDS)، و همچنین مباحث مرتبط با ایندکس‌گذاری و جست‌وجوی داده‌ها هستند. شناخت ویژگی‌ها و ساختار داده‌های تصویربرداری مغز، پایه‌ای برای انتخاب فرمت و استاندارد مناسب ذخیره‌سازی است. این انتخاب هم تعیین می‌کند که از چه ابزارها و روش‌هایی برای پردازش، جست‌وجو و مدیریت داده‌ها می‌توان استفاده کرد. در ادامه مباحث این فصل، به بررسی و تحلیل کارهای مرتبط پرداخته شده تا ضمن شناسایی راهکارهای موجود، نقاط قوت و ضعف آن‌ها مشخص و مسیر پیش روی پروژه حاضر ترسیم شود.

۲-۲ مفاهیم پایه

۱-۲-۲ داده‌های تصویربرداری مغز و انواع آن

تصویربرداری عصبی یک اصطلاح کلی برای مجموعه‌ای از روش‌ها و فناوری‌های غیرتهاجمی^{۱۱} است که به منظور بررسی ساختار، عملکرد و ارتباطات عصبی مغز استفاده می‌شوند. این روش‌ها داده‌های متنوعی از جنبه‌های ساختاری و عملکردی مغز ارائه می‌دهند و نقش مهمی در علوم اعصاب و پزشکی مدرن ایفا می‌کنند. در واقع می‌توان گفت هر روش تصویربرداری، زاویه و بُعد متفاوتی از مغز را نشان می‌دهد و بر اساس نیاز پژوهش یا تشخیص، می‌توان از یک یا چند روش به صورت مجزا یا ترکیبی استفاده کرد.

از میان روش‌های ساختاری، تصویربرداری تشدید مغناطیسی ساختاری^{۱۲} تصاویر با وضوح مکانی بسیار بالا از بافت‌های نرم مغز ارائه می‌دهد و برای تشخیص ضایعات یا تغییرات آناتومیکی بسیار ارزشمند است. همچنین، تصویربرداری تانسور انتشار^{۱۳} مسیرهای ماده سفید مغز را مشخص می‌کند و برای مطالعه ارتباطات عصبی به کار

¹¹ Non-invasive

¹² Structural Magnetic Resonance Imaging (sMRI)

¹³ Diffusion Tensor Imaging (DTI)

می‌رود. روش‌هایی مانند توموگرافی کامپیوتری^{۱۴} نیز اطلاعات ساختاری مفیدی فراهم می‌کنند. از طرف دیگر در حوزه روش‌های عملکردی، تصویربرداری تشدید مغناطیسی عملکردی^{۱۵} با اندازه‌گیری تغییرات جریان خون، فعالیت بخش‌های مختلف مغز را هنگام انجام وظایف مشخص (شناختی یا حرکتی) ثبت می‌کند. از طرف دیگر تکنیک‌های الکتروفیزیولوژیک مانند الکتروانسفالوگرافی^{۱۶} و مگنتوانسفالوگرافی^{۱۷}، فعالیت الکتریکی یا میدان‌های مغناطیسی ناشی از فعالیت نورونی را ثبت می‌کنند. علاوه بر این، روش‌هایی مانند توموگرافی انتشار پوزیترون^{۱۸} و طیف‌سنجی نزدیک به مادون قرمز^{۱۹} نیز داده‌های عملکردی ارزشمندی فراهم می‌کنند [۶][۷].

هر یک از این تکنیک‌ها مزایا و معایبی مربوط به وضوح، ایمنی و دسترسی‌پذیری دارند. به عنوان مثال، EEG وضوح زمانی بالا اما وضوح مکانی پایین دارد، در حالی که fMRI وضوح مکانی بالا اما وضوح زمانی پایینی دارد [۶]؛ بنابراین انتخاب روش مناسب به هدف پژوهش یا تشخیص بستگی دارد. در بسیاری از موارد، ترکیب دو یا چند روش (مانند fMRI و EEG/MEG) می‌تواند مزایای هر یک را در کنار هم فراهم کرده و دید جامع‌تری از فعالیت و ساختار مغز ارائه دهد.

۲-۱-۱-۲-۲ داده‌های چندوجهی در تصویربرداری مغزی

داده‌های چندوجهی زمانی به دست می‌آیند که اطلاعات مغز از چندین روش مختلف گردآوری شود. هر روش تصویربرداری دیدگاه و محدودیت خاص خود را دارد و به‌تنهایی قادر به ارائه تصویر جامع از مغز نیست؛ اما ترکیب آن‌ها نمایی چندبعدی و دقیق‌تر از ساختار و عملکرد مغز فراهم می‌کند. این رویکرد نه‌تنها دقت تشخیص بیماری‌ها را افزایش می‌دهد، بلکه درک فرآیندهای پردازش اطلاعات در مغز را عمیق‌تر کرده و داده‌های ارزشمندی برای مدل‌های هوش مصنوعی و یادگیری ماشین فراهم می‌آورد.

با وجود این مزایا، پردازش داده‌های چندوجهی با چالش‌هایی روبه‌روست. هم‌ترازسازی^{۲۰} داده‌ها به دلیل تفاوت وضوح، مقیاس و زاویه در روش‌های مختلف نیازمند الگوریتم‌های پیشرفته است. از طرف دیگر حجم بالای داده‌ها نیز به ظرفیت بالای ذخیره‌سازی و پردازش نیاز دارد. همچنین، ناهمگونی ذاتی بین رزولوشن زمانی و مکانی

¹⁴ Computed Tomography (CT)

¹⁵ Functional Magnetic Resonance Imaging (fMRI)

¹⁶ Electroencephalography (EEG)

¹⁷ Magnetoencephalography (MEG)

¹⁸ Positron Emission Tomography (PET)

¹⁹ Near-Infrared Spectroscopy (NIRS)

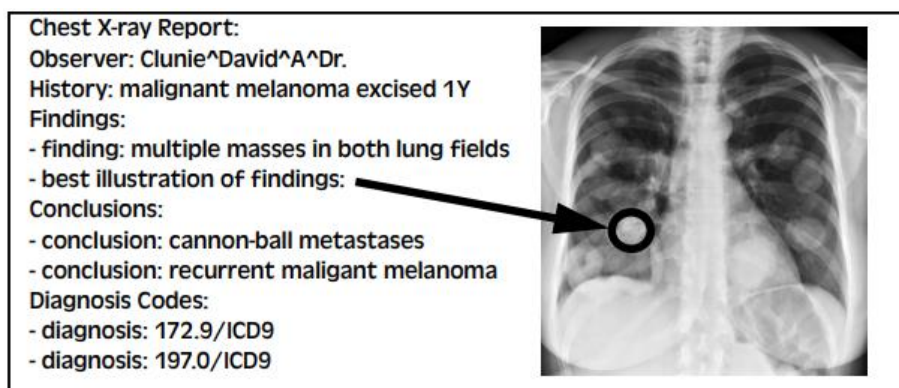
²⁰ Coregistration

روش‌هایی مانند EEG و fMRI ادغام را پیچیده‌تر می‌کند [۶]. نبود استانداردهای یکپارچه برای قالب‌بندی و ثبت داده‌ها نیز مانعی برای تبادل و استفاده مجدد آن‌هاست، هرچند چارچوب‌هایی مانند BIDS گام‌های مهمی در رفع این مشکل برداشته‌اند [۴].

۲-۲-۲ فرمت‌ها و استانداردهای ذخیره‌سازی داده‌ها

۱-۲-۲-۲ فرمت DICOM

فرمت DICOM استاندارد بین‌المللی برای ذخیره‌سازی، تبادل و مدیریت داده‌های تصویربرداری پزشکی است که علاوه بر خود تصویر، فراداده مربوط به آن مانند نوع دستگاه، تاریخ تصویربرداری، مشخصات بیمار، پارامترهای فنی و... را نیز در یک فایل یکپارچه نگه می‌دارد. DICOM امکان تبادل ایمن و سازگار تصاویر پزشکی بین دستگاه‌ها، نرم‌افزارها و مراکز درمانی مختلف را فراهم می‌کند و از بروز ناسازگاری بین سیستم‌ها جلوگیری می‌نماید. علاوه بر تصاویر خام، DICOM قابلیت پشتیبانی از اشیاء گزارش ساختاریافته^{۲۱} را دارد که می‌توانند شامل توضیحات، یادداشت‌ها یا ارجاعات به تصاویر و سایر اشیاء باشند [۸].



شکل ۲- نمونه‌ای ساده از یک ساختار گزارش DICOM

²¹ Structured Report (SR)

۲-۲-۲-۲ فرمت NIfTI

فرمت NIfTI یک استاندارد رایج برای ذخیره و تبادل داده‌های تصویربرداری مغزی، به‌ویژه MRI، است که با هدف ساده‌سازی پردازش و تحلیل داده‌ها توسعه یافته است. NIfTI برخلاف DICOM که بیشتر برای تبادل و آرشیو بالینی استفاده می‌شود، ساختار آن برای پردازش‌های محاسباتی و نرم‌افزاری بهینه شده و توسط بسیاری از ابزارهای تحلیلی پشتیبانی می‌شود. این فرمت امکان ذخیره داده‌های حجمی سه‌بعدی و چهاربعدی (مانند سری‌های زمانی fMRI) را فراهم کرده و از یک هدر شامل اطلاعات کلیدی مانند ابعاد، رزولوشن، نوع داده و سیستم مختصات استفاده می‌کند. NiFTI-۱ می‌تواند به صورت یک فایل واحد (.nii) یا دو فایل جدا (.img و .hdr) ذخیره شود [۹].

۲-۲-۲-۳ استاندارد BIDS

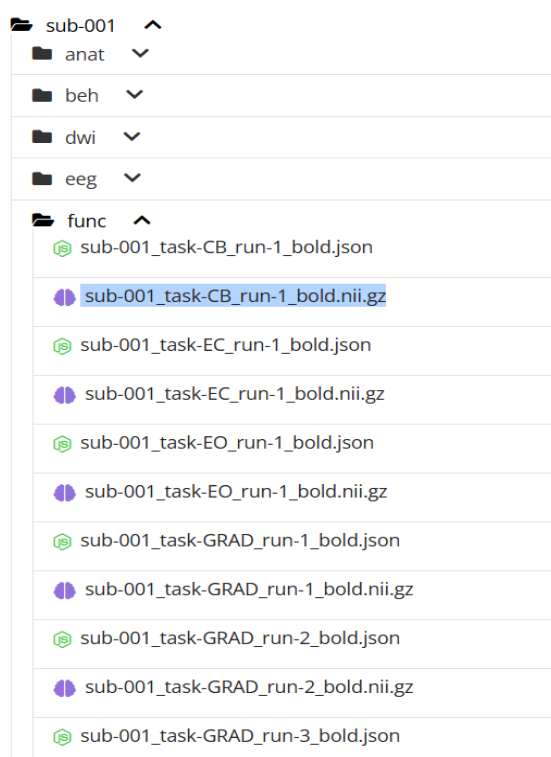
BIDS یک استاندارد برای سازماندهی و توصیف داده‌های تصویربرداری مغزی است که با هدف رفع مشکل پراکندگی و نبود توافق عمومی در ساختاردهی داده‌های MRI و دیگر داده‌های علوم اعصاب ایجاد شده است. پیش از معرفی BIDS، هر آزمایشگاه یا پژوهشگر معمولاً داده‌های خود را به صورت منحصر به فرد و گاهی غیرسازگار با نرم‌افزارها و خطوط پردازش رایج ذخیره می‌کرد، که این امر اشتراک‌گذاری، استفاده مجدد، و اجرای خودکار فرایندهای تحلیل را دشوار یا حتی غیرممکن می‌ساخت [۱۰]. BIDS با تعریف یک ساختار پوشه‌ای منظم و استفاده از فرمت‌های فایل رایج و سازگار مانند NIfTI برای داده‌های تصویری، JSON برای فراداده‌ها و TSV برای جداول متنی، رویکردی ساده، شهودی و درعین حال قدرتمند ارائه می‌دهد که هم با رویه‌های عملی آزمایشگاهی همخوان است و هم برای پژوهشگرانی با پیشینه فنی محدود قابل استفاده است [۴].

BIDS با یکسان‌سازی نحوه ذخیره‌سازی داده‌ها و فراداده‌ها، امکان تبادل آسان‌تر داده‌ها میان پژوهشگران، ادغام مجموعه داده‌های مختلف، اجرای خطوط پردازش خودکار و کاهش خطاهای ناشی از ناسازگاری ساختارها را فراهم می‌کند. این استاندارد همچنین پشتیبانی وسیعی از انواع داده‌ها، از جمله MRI ساختاری و عملکردی، DTI، EEG و MEG دارد. به این ترتیب، BIDS نه تنها مدیریت و تحلیل داده‌های پیچیده تصویربرداری مغزی را ساده‌تر می‌کند، بلکه با افزایش شفافیت علمی، به ارتقاء کیفیت و سرعت تحقیقات علوم اعصاب کمک شایانی می‌کند.

۱-۳-۲-۲-۲ ساختار سلسله‌مراتبی BIDS

در استاندارد BIDS، ساختار فایل‌ها به صورت سلسله‌مراتبی و سازمان‌یافته طراحی شده است تا داده‌ها به طور قابل‌فهم ذخیره شوند. در بالاترین سطح، هر شرکت‌کننده با یک پوشه‌ی $sub- < subject_id >$ مشخص می‌شود. در سطح بعدی برای هر فرد، داده‌ها بر اساس مدالیته (مانند anat برای MRI ساختاری، func برای fMRI عملکردی و...) دسته‌بندی می‌شوند. همچنین نام‌گذاری فایل‌ها ترکیبی از برچسب‌های معنادار است، مانند شناسه‌ی فرد، نوع وظیفه^{۲۲}، شماره اجرای تکرار^{۲۳} و نوع داده [۵].

برای مثال فایل "sub-001_task-CB_run-1_bold.nii.gz" در شکل ۳، یک فایل fMRI است که به آزمودنی ۰۰۱، وظیفه الگوی شطرنجی^{۲۴}، و اجرای اول مربوط می‌شود.



شکل ۳- ساختار یک مجموعه داده سازگار با BIDS

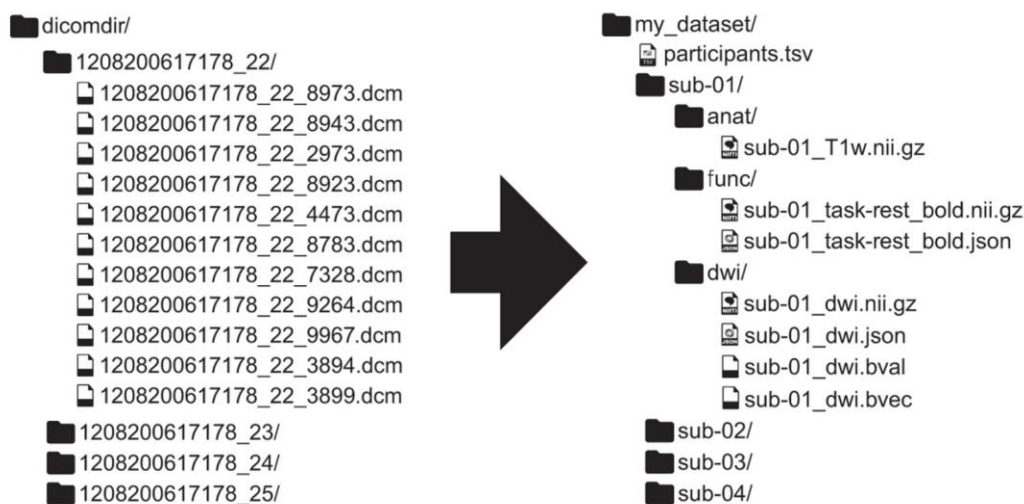
²² Task

²³ Run

²⁴ CheckerBoard

۲-۲-۲-۲-۲ فرایند تبدیل به BIDS

در فرایند تبدیل داده‌های MRI به قالب BIDS، نخست داده‌ها مستقیماً از اسکنر MRI دریافت می‌شوند که معمولاً در قالب DICOM ذخیره شده‌اند. همانطور که قبل‌تر اشاره کردیم، این فرمت شامل تصویر خام به همراه فراداده‌های فنی و اطلاعات بیمار است. در گام اول، این داده‌ها با استفاده از ابزارهایی مانند dcm2nii به فرمت NIfTI همراه با فایل‌های JSON حاوی فراداده‌های کلیدی تبدیل می‌شوند. این مرحله همچنین ناسازگاری‌های ناشی از تفاوت در پیاده‌سازی DICOM توسط تولیدکنندگان مختلف اسکنر را تا حد زیادی برطرف می‌کند [۵].

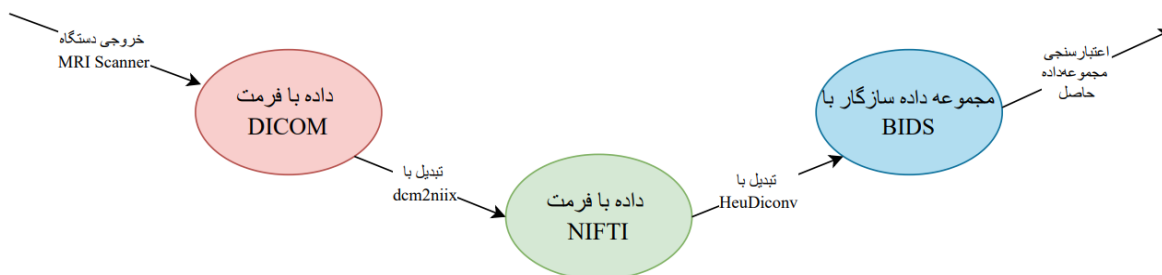


شکل ۴- خروجی‌های آزمایش‌های تصویربرداری عصبی (چپ) و تبدیل آن به مجموعه داده سازگار با BIDS

در گام بعد، فایل‌های NIfTI و فراداده بر اساس ساختار پوشه‌ها و نام‌گذاری استاندارد BIDS سازمان‌دهی می‌شوند. ابزارهایی مانند HeuDicomv این کار را با بهره‌گیری از اسکریپت‌های ابتکاری^{۲۵} انجام می‌دهند و داده‌ها را در پوشه‌های مرتبط با هر آزمودنی و هر نوع اسکن قرار می‌دهند. در ادامه، داده‌های کمکی مانند فایل‌های اطلاعات دموگرافیک شرکت‌کنندگان اضافه می‌شوند. همچنین، فایل‌های فراداده سطح مطالعه مانند dataset_description.json برای توصیف کلی داده‌ها ایجاد می‌گردد. در پایان، مجموعه داده با استفاده از

²⁵ Heuristic

اعتبارسنج^{۲۶} بررسی می‌شود تا از رعایت کامل استاندارد و کامل بودن فراداده اطمینان حاصل شود [۱۱]. سیر تکاملی ذکر شده در شکل ۵ قابل مشاهده است.



شکل ۵- فرایند تبدیل داده خام به قالب BIDS

۳-۲-۲ ابزارها و کتابخانه‌های مرتبط با BIDS

چند ابزار کلیدی در اکوسیستم BIDS وجود دارند که فرآیند تبدیل، سازمان‌دهی و تحلیل داده‌های تصویربرداری عصبی را ساده و استاندارد می‌کنند. این ابزارها در کنار هم، چرخه کامل کار با BIDS را پوشش می‌دهند: dcm2niix برای تبدیل، HeuDiConv برای سازمان‌دهی و ساختاردهی، و PyBIDS برای جست‌وجو، پردازش و تحلیل داده‌های نهایی.

۳-۲-۲-۱ ابزار dcm2niix

dcm2niix یکی از پرکاربردترین ابزارها برای تبدیل داده‌های تصویربرداری پزشکی از قالب DICOM (خروجی رایج اسکنرهای MRI) به فرمت NIfTI است. این ابزار علاوه بر ایجاد فایل NIfTI، یک فایل جانبی JSON سازگار با BIDS نیز تولید می‌کند که شامل فراداده مهم و استانداردشده است. سادگی و سرعت بالا، پشتیبانی چندسکویی^{۲۷} و رفع ناسازگاری‌های ناشی از تفاوت پیاده‌سازی DICOM توسط سازندگان مختلف از ویژگی‌های آن است [۱۳].

^{۲۶} BIDS Validator

^{۲۷} Cross-platform

۲-۳-۲-۲ ابزار HeuDiConv

HeuDiConv ابزاری انعطاف‌پذیر برای تبدیل و سازمان‌دهی داده‌های DICOM به ساختار پوشه‌ها و نام‌گذاری استاندارد BIDS است. این ابزار با استفاده از dcm2niix برای تبدیل داده‌ها، و با بهره‌گیری از اسکریپت‌های ابتکاری، امکان تعریف الگوهای دلخواه برای سازمان‌دهی فایل‌ها را فراهم می‌کند. قابلیت‌هایی مانند ثبت مسیر تبدیل در قالب W3C PROV، یکپارچگی با DataLad برای مدیریت نسخه و داده، و پشتیبانی از درج خودکار در خط لوله^{۲۸}های دریافت داده از مزایای آن است [۱۱].

۳-۳-۲-۲ کتابخانه PyBIDS

PyBIDS یک کتابخانه پایتونی است که کار با داده‌های BIDS را برای پژوهشگران و توسعه‌دهندگان ساده می‌کند. این کتابخانه امکاناتی مانند جست‌وجو بر اساس فراداده یا کلمات کلیدی، بازیابی و تجمیع فراداده پراکنده در لایه‌های مختلف ساختار BIDS، ساخت مسیرهای معتبر برای فایل‌های جدید، و اعتبارسنجی داده‌ها بر اساس مشخصات BIDS را ارائه می‌دهد. PyBIDS از طریق pip به‌سادگی نصب می‌شود و تعامل با آن عمدتاً از طریق شیء BIDSLayout انجام می‌گیرد [۱۲][۱۴]. برای مثال کد موجود در شکل ۶ همه فراداده‌های فایل‌های یک مجموعه داده BIDS را استخراج کرده، مسیر هر فایل را اضافه می‌کند و نتیجه را در یک فایل JSON ذخیره می‌کند.

```
import json
from bids import BIDSLayout

layout = BIDSLayout(r"BIDS\datasets\ds006012")
data = [{"**layout.get_metadata(f.path), "path": f.path} for f in layout.get()]

with open("bids_metadata.json", "w") as f:
    json.dump(data, f, indent=4)
```

شکل ۶- استخراج فراداده با کتابخانه PyBIDS

²⁸ Pipeline

۲-۲-۴ ایندکس گذاری و جست و جو در داده ها

۲-۲-۴-۱ مفهوم ایندکس گذاری

ایندکس گذاری فرایندی است که طی آن داده ها به شکلی سازمان یافته و بهینه ذخیره می شوند تا بتوان آن ها را سریع و کارآمد بازیابی کرد. به جای جست و جوی مستقیم در کل مجموعه داده که می تواند زمان بر و پرهزینه باشد، اطلاعات کلیدی هر رکورد در ساختارهایی مانند جداول هش یا درخت های جست و جو ذخیره می شود. این ساختارها که ایندکس نام دارند، همانند فهرست یک کتاب، به سیستم کمک می کنند تا محل دقیق داده های مورد نظر را با کمترین زمان و منابع بیابد. در حوزه داده های علمی و حجیم، ایندکس گذاری هم سرعت جست و جو را بهبود می بخشد، هم امکان انجام پرس و جو^{۲۹} های پیچیده و فیلترگذاری چندمرحله ای را نیز فراهم می کند.

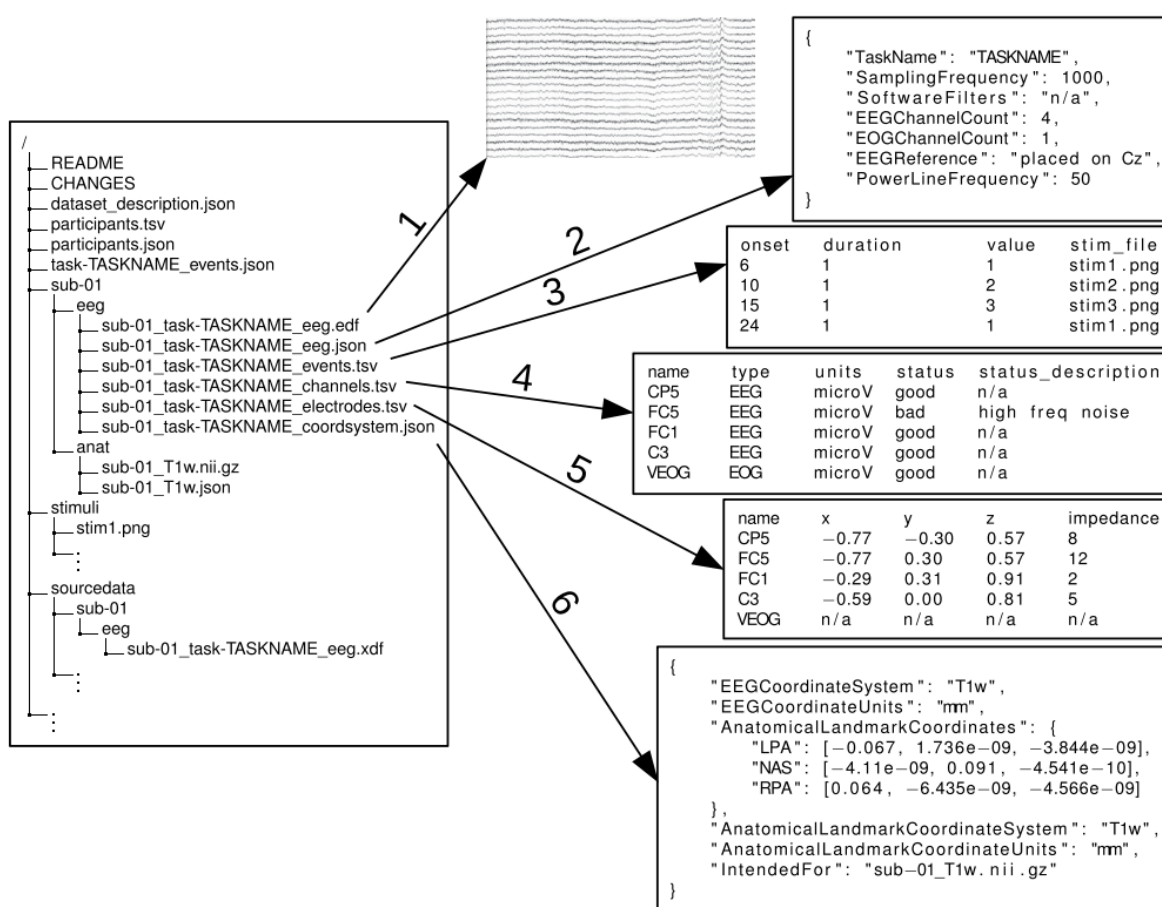
۲-۲-۴-۲ موتور جستجوی Elasticsearch

Elasticsearch یک موتور جست و جوی متن باز است که برای ایندکس گذاری و جست و جوی سریع داده ها به ویژه داده های متنی و ساخت یافته در قالب JSON به کار می رود. این ابزار قادر است داده ها را به صورت اسناد ذخیره و آن ها را در شاخص های بهینه سازی شده سازمان دهی کند، به گونه ای که جست و جو حتی در مجموعه های داده چند ترابایتی نیز تقریباً به صورت آنی انجام شود. در این پروژه، از Elasticsearch به دلیل توانایی بالای آن در مدیریت داده های حجیم، پشتیبانی از جست و جوهای پیشرفته مانند جست و جوی مبتنی بر شباهت یا فیلترگذاری چندلایه و مقیاس پذیری بالا استفاده کرده ایم تا بتوان داده های تصویربرداری مغزی و فراداده های مرتبط را به سرعت ایندکس، بازیابی و تحلیل کرد.

²⁹ Query

۳-۲ تحلیل نقاط قوت و ضعف کارهای مشابه

استانداردسازی داده‌های تصویربرداری مغزی طی سال‌های اخیر با معرفی BIDS دچار تحول اساسی شد. این استاندارد که نخستین بار توسط *Gorgolewski* و همکارانش معرفی گردید، چارچوبی مشخص برای سازماندهی، توصیف و اشتراک‌گذاری داده‌های MRI و سایر مدالیت‌ها ارائه داد و باعث شد پردازش خودکار و تضمین کیفیت داده‌ها تسهیل شود [۴]. در ادامه، توسعه‌های مختلفی تحت عنوان BIDS extensions برای پشتیبانی از سایر روش‌های تصویربرداری مانند EEG، MEG و PET ارائه گردید [۱۶].



شکل ۷- تشریح نحوه پشتیبانی از مدالیت EEG در نسخه جدید BIDS

برای آماده‌سازی داده‌ها در قالب BIDS، ابزارهای گوناگونی توسعه یافته‌اند. *dcm2niix* به‌عنوان یکی از پرکاربردترین مبدل‌ها، داده‌های خام DICOM را به فرمت NIfTI به‌همراه فایل‌های جانبی JSON تبدیل می‌کند.

و بسیاری از تفاوت‌های ناشی از پیاده‌سازی‌های متفاوت سازندگان اسکنر را نرمال‌سازی می‌نماید [۱۳]. ابزار HeuDiConv نیز به‌عنوان چارچوبی انعطاف‌پذیر بر پایه dcm2niix توسعه یافته و امکان تعریف هیوریستیک‌ها برای تعیین خودکار نام‌گذاری فایل‌ها، ساختاردهی پوشه‌ها و حتی نگهداری پیوند به داده‌های اصلی را فراهم می‌آورد که در پروژه‌های بزرگ باعث صرفه‌جویی قابل توجهی در زمان می‌شود. [۱۱]. در ادامه، بسته PyBIDS نقش مهمی در استخراج کارآمد داده‌های BIDS، بازیابی و یکپارچه‌سازی فراداده و حتی پشتیبانی از ساخت مدل‌های آماری مطابق با BIDS-StatsModel ایفا می‌کند [۱۲].

در کنار این ابزارها، اعتبارسنج ساختار داده‌های تصویربرداری مغز^{۳۰} به‌عنوان یک وب‌اپلیکیشن، ابزار خط فرمان و کتابخانه جاوااسکریپت/تایپاسکریپت^{۳۱} توسعه یافته تا انطباق داده‌ها با استاندارد BIDS را ارزیابی کند. نسخه وب این ابزار به‌صورت درون‌مرورگری^{۳۲} اجرا می‌شود و هیچ داده‌ای را به سرور خارجی منتقل نمی‌کند، که آن را برای کار با داده‌های حساس ایمن می‌سازد. همچنین در شرایطی مانند کار بر روی سرورهای راه دور، نسخه خط فرمان آن امکان اعتبارسنجی سریع و بدون وابستگی به محیط گرافیکی را فراهم می‌کند. این ابزار به‌عنوان بخشی از چرخه آماده‌سازی داده‌ها در بسیاری از پروژه‌های علوم اعصاب محاسباتی مورد استفاده قرار می‌گیرد [۱۵].

با وجود شباهت‌ها، هر یک از این ابزارها رویکرد و محدودیت‌های خاص خود را دارند. برای مثال، dcm2niix در عین سادگی و سرعت بالا، انعطاف کمی در سازمان‌دهی پوشه‌ها و نام‌گذاری فایل‌ها دارد و معمولاً در پروژه‌های کوچک یا در مراحل اولیه آماده‌سازی داده به کار می‌رود. در مقابل، HeuDiConv با بهره‌گیری از هیوریستیک‌ها و امکان پیوند به داده‌های اصلی، گزینه‌ای مناسب برای پروژه‌های بزرگ و بایوبانک‌ها محسوب می‌شود، در واقع می‌توان گفت HeuDiConv با بهره‌گیری از dcm2niix، فرایند تبدیل داده خام به قالب BIDS را یکپارچه کرده است. PyBIDS نیز اگرچه قابلیت‌های قدرتمندی برای جستجو و استخراج فراداده دارد، اما به‌تنهایی فرایند آماده‌سازی اولیه داده‌ها را انجام نمی‌دهد و بیشتر به‌عنوان لایه‌ای مکمل پس از سازمان‌دهی داده‌ها در قالب BIDS عمل می‌کند. در این پروژه نیز از PyBIDS برای استخراج و یکپارچه‌سازی فراداده استفاده شد، اما با توجه به حجم بالای داده‌ها و نیاز به سرعت در جستجو و بازیابی، بار اصلی پردازش‌های جستجو به موتور Elasticsearch سپرده شد تا امکان پاسخ‌گویی سریع و مقیاس‌پذیر فراهم شود.

در جریان مطالعه این ابزارها، مشخص شد که هرچند HeuDiConv، dcm2niix و BIDS Validator نقش مهمی در آماده‌سازی و اعتبارسنجی داده‌ها دارند، اما تمرکز این پروژه بر مرحله‌ای پس از آماده‌سازی اولیه بوده

³⁰ BIDS Validator

³¹ JavaScript/TypeScript

³² Client-side

است. از آنجا که داده‌ها از ابتدا در قالب BIDS در دسترس بودند، نیاز به انجام فرآیند تبدیل یا اعتبارسنجی اولیه وجود نداشت. بنابراین، تمرکز کار بر روی طراحی سازوکاری برای مدیریت، ایندکس‌گذاری و بازیابی کارآمد داده‌ها و فراداده قرار گرفت؛ رویکردی که به‌ویژه در مقیاس بایوبانک‌ها اهمیت دارد، زیرا اضافه‌شدن مستمر داده‌های جدید و تنوع مدالیت‌ها نیازمند سیستمی پویا و مقیاس‌پذیر است.

۴-۲ جمع‌بندی

در این فصل، ابتدا مفاهیم و ابزارهای کلیدی مرتبط با داده‌های تصویربرداری مغزی و استانداردهای ذخیره‌سازی آن‌ها معرفی و تحلیل شد و سپس، نمونه‌های برجسته از پروژه‌ها و سامانه‌های مشابه مورد بررسی قرار گرفت تا جایگاه و مسیر توسعه پروژه حاضر روشن‌تر شود. این مرور، زمینه‌ای برای طراحی رویکرد پیشنهادی فراهم کرده است. در فصل سوم، بر مبنای این چارچوب، روش تحقیق و طراحی سیستم پیشنهادی تشریح خواهد شد که شامل معماری کلی، فرایندهای پیاده‌سازی، ابزارهای مورد استفاده، و جزئیات فنی مراحل انجام پروژه است.

فصل سوم: روش پیشنهادی و نتیجه‌گیری

۳-۱ مقدمه

در این فصل، فرآیند طراحی و پیاده‌سازی روش پیشنهادی برای ایجاد یک سامانه‌ی یکپارچه در مدیریت و تحلیل داده‌های تصویربرداری مغزی مبتنی بر استاندارد BIDS تشریح می‌شود. هدف اصلی، ارائه‌ی معماری‌ای است که بتواند داده‌ها و فراداده‌های پیچیده‌ی حاصل از تصویربرداری‌های چندوجهی را به‌صورت ساخت‌یافته ذخیره، ایندکس و بازیابی کرده و در عین حال امکان پایش و مصورسازی آن‌ها را نیز فراهم آورد.

برای دستیابی به این هدف، ابتدا نیازمندی‌های سیستم و معماری کلی آن معرفی می‌شود. سپس مراحل اجرایی پروژه به‌صورت فازبندی شده توضیح داده خواهد شد و نحوه‌ی جریان داده‌ها در سامانه تبیین می‌گردد. در ادامه، اجزای مختلف پیاده‌سازی شامل ماژول‌های پردازش و ایندکس‌گذاری، پایش و مصورسازی، رابط کاربری، به‌روزرسانی خودکار با Airflow و یکپارچه‌سازی زیرساخت با داکر به تفصیل بیان می‌شوند. در نهایت، روش ارزیابی و سناریوهای طراحی‌شده برای سنجش کارایی سامانه مورد بحث قرار خواهند گرفت.

۳-۲ ساختار روش پیشنهادی

هدف اصلی این پروژه طراحی و پیاده‌سازی یک زیرساخت مدیریت داده برای بایوبانک داده‌های مغزی در قالب استاندارد BIDS است که بتواند حجم بالای داده‌های چندوجهی^{۳۳} را با کارایی مناسب ذخیره، ایندکس‌گذاری، جستجو و پایش کند. ایده اصلی این است که داده‌های ورودی (اعم از داده‌های تصویری، فراداده‌ها و داده‌های رفتاری) پس از ورود به سیستم، ابتدا استخراج و پردازش می‌شوند، سپس در قالب ایندکس‌های مجزا در Elasticsearch ذخیره شده و در نهایت با استفاده از Kibana به شکل داشبوردهای تحلیلی در دسترس پژوهشگر قرار می‌گیرند. همچنین یک رابط کاربری تحت وب با Flask طراحی شده که امکان افزودن داده‌های جدید به سیستم را فراهم می‌سازد. به‌منظور خودکارسازی عملیات، فرآیندهای ایندکس‌گذاری و به‌روزرسانی داده‌ها از طریق گردش کارهای Airflow زمان‌بندی و اجرا می‌شوند.

³³ Multi-modal

۳-۲-۱ نیازمندی‌های سیستم

سیستم پیشنهادی با توجه به نیازهای بایوبانک داده‌های مغزی، نیازمندی‌های زیر را پوشش می‌دهد که این نیازمندی‌ها به دو بخش عملکردی^{۳۴} و غیرعملکردی^{۳۵} تقسیم می‌شوند:

● نیازمندی‌های عملکردی:

- امکان جستجو در مجموعه داده
- امکان فیلتر کردن دادگان بر حسب ویژگی‌های مختلف
- امکان افزودن آزمودنی جدید
- مصورسازی دادگان و فراداده‌های مجموعه داده و ارائه داشبوردهای تصویری برای پایش و تحلیل اطلاعات کلیدی
- به‌روزر بودن داشبوردها با تغییر مجموعه داده

● نیازمندی‌های غیرعملکردی:

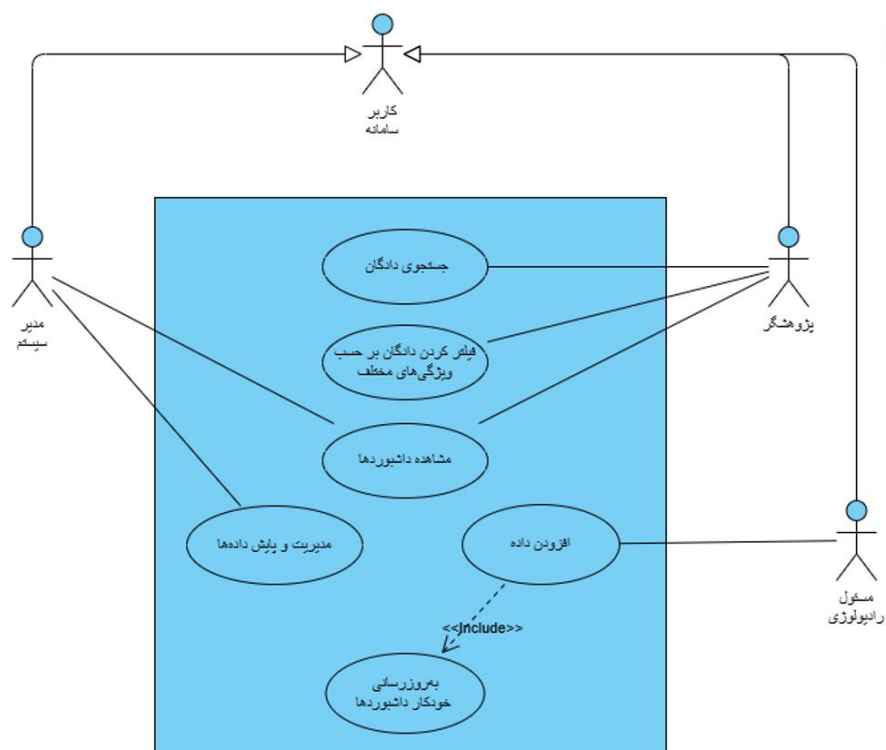
- جستجوی سریع و انعطاف‌پذیر روی داده‌ها و فراداده‌ها
- امکان پاسخگویی سیستم در صورت بزرگ شدن مجموعه داده
- پایداری سیستم در طول زمان

در شکل ۸ نمودار موارد کاربردی^{۳۶} سیستم پیشنهادی نشان داده شده است که تعامل سه نقش اصلی با سامانه را نمایش می‌دهد. پژوهشگر می‌تواند داده‌ها را جستجو و فیلتر کرده و داشبوردهای تحلیلی را مشاهده کند؛ مدیر سیستم وظیفه مدیریت و پایش داده‌ها را بر عهده دارد؛ و مسئول رادیولوژی از طریق رابط کاربری وب امکان افزودن داده‌های جدید را دارد.

³⁴ Functional

³⁵ Non-functional

³⁶ Use Case Diagram



شکل ۸- نمودار موارد کاربری

۳-۲-۲ نمای کلی سیستم و معماری پیشنهادی

سیستم پیشنهادی در پنج لایه اصلی طراحی شده است که هر کدام وظیفه‌ی مشخصی در مدیریت، پردازش و تحلیل داده‌ها برعهده دارند. این معماری در شکل ۹ قابل مشاهده است. این لایه‌ها عبارت‌اند از:

۱. لایه دریافت داده^{۳۷}:

در این لایه، داده‌های جدید از طریق رابط کاربری مبتنی بر Flask دریافت و به مجموعه‌داده فعلی افزوده می‌شوند. هر بسته‌ی داده‌ای به صورت فایل فشرده^{۳۸} بارگذاری شده و در ساختار استاندارد BIDS قرار می‌گیرد

³⁷ Data Ingestion

³⁸ ZIP

و فراداده مربوط به آن در فایل TSV ثبت می‌گردد تا امکان جستجو و پردازش دقیق‌تر فراهم شود. این لایه به‌عنوان نقطه‌ی ورودی سیستم عمل می‌کند و وظیفه‌ی سازماندهی اولیه داده‌ها را بر عهده دارد.

۲. لایه پردازش^{۳۹}:

در این لایه، اسکریپت‌های پایتونی داده‌های ورودی را پردازش کرده و اطلاعات مورد نیاز از فایل‌های تصویری، فراداده‌های JSON و TSV و سایر اجزای موجود در ساختار BIDS استخراج می‌شود. خروجی این لایه شامل داده‌های ساخت‌یافته در قالب رکوردهای استاندارد است که آماده ذخیره‌سازی در پایگاه جستجو می‌باشند.

۳. لایه ایندکس‌گذاری^{۴۰}:

داده‌های پردازش‌شده در این مرحله به سامانه‌ی Elasticsearch منتقل می‌شوند. برای هر دسته‌ی اطلاعات شامل آزمودنی‌ها، فایل‌ها و دادگان آزمایشات، ایندکس‌های جداگانه ایجاد می‌گردد. این ساختار ایندکس‌گذاری امکان جستجوی سریع و بازیابی مؤثر داده‌ها را فراهم می‌سازد و زیرساخت اصلی برای عملیات پایش و تحلیل در لایه‌های بالاتر محسوب می‌شود.

۴. لایه زمان‌بندی^{۴۱} و به‌روزرسانی:

برای تضمین اجرای منظم و خودکار پردازش‌ها، هر اسکریپت در قالب یک جریان کاری^{۴۲} در سامانه‌ی Airflow تعریف شده است. این لایه به صورت موازی با سایر لایه‌ها اجرا می‌شود وظیفه‌ی زمان‌بندی، کنترل و نظارت بر فرآیندهای پردازش و ایندکس‌گذاری را بر عهده دارد و از به‌روزرسانی مستمر و روزانه پایگاه داده اطمینان حاصل می‌کند.

۵. لایه پایش^{۴۳} و مصورسازی^{۴۴}:

در بالاترین سطح، داشبوردهای Kibana بر اساس ایندکس‌های ایجاد شده طراحی شده‌اند. این داشبوردها امکان مشاهده، تحلیل و پایش داده‌ها را برای کاربران فراهم می‌کنند و با استفاده از انواع نمودارها و شاخص‌های

³⁹ Process

⁴⁰ Indexing

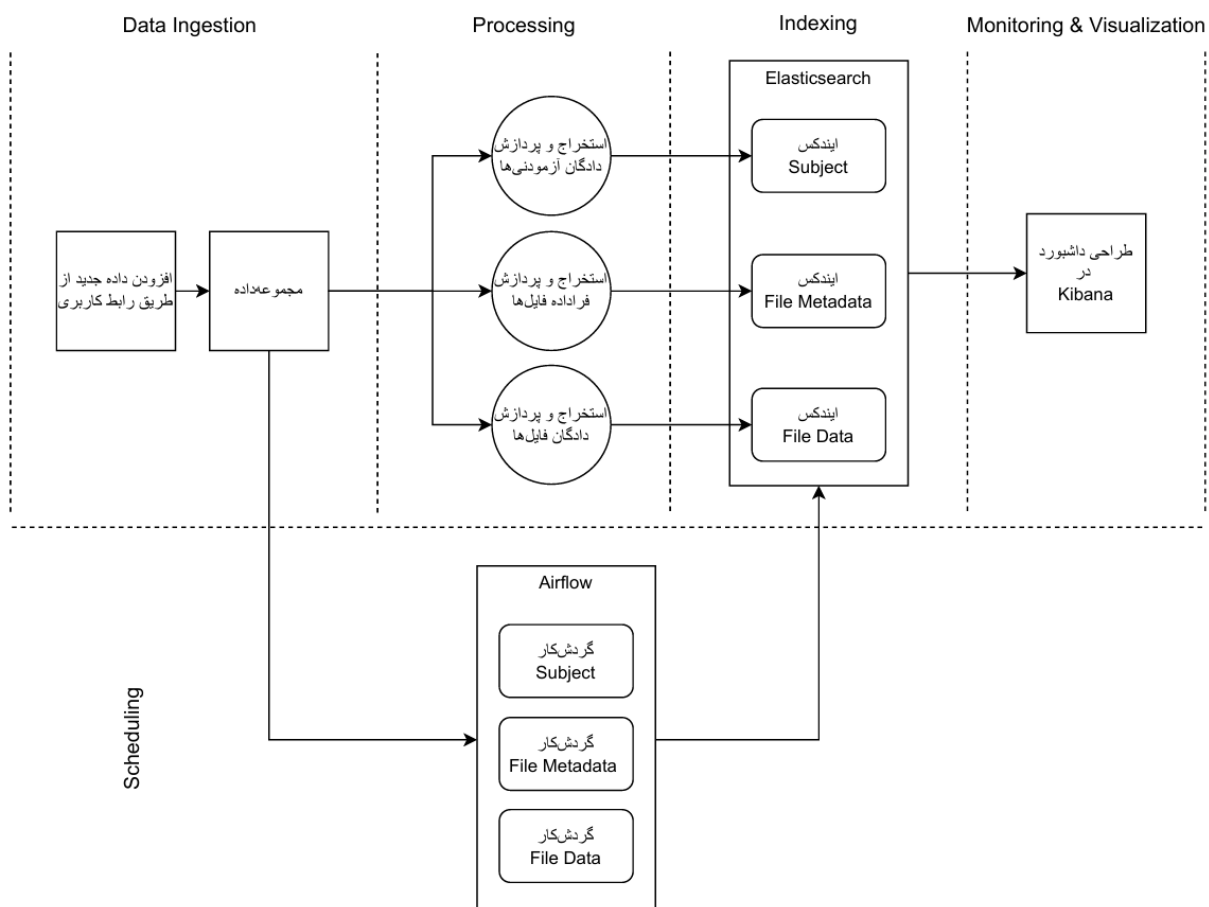
⁴¹ Scheduling

⁴² DAG

⁴³ Monitoring

⁴⁴ Visualization

تحلیلی، تصویری جامع از وضعیت داده‌ها و کیفیت آن‌ها ارائه می‌دهند. بدین ترتیب، پژوهشگران و مدیران می‌توانند با یک نگاه کلی، وضعیت بایوبانک و داده‌های موجود را ارزیابی کنند.



شکل ۹- نمای معماری کلان سیستم

۳-۲-۳ فازبندی اجرای پروژه

به منظور مدیریت ساختارمند پروژه و دستیابی تدریجی به اهداف، کل فرآیند در چند فاز متوالی تعریف و اجرا شده است.

فاز اول: تحقیق و مطالعه ابزارهای مورد نیاز

در فاز آغازین به مطالعه و بررسی منابع موجود در حوزه‌ی BIDS پرداختیم. تمرکز اصلی بر شناخت این استاندارد و ابزارهای پرکاربرد در این حوزه بود. به همین منظور، ساختار BIDS و کتابخانه‌ی PyBIDS به عنوان مبنای پردازش داده‌ها مورد بررسی قرار گرفتند. در ادامه، قابلیت‌های سامانه‌ی Elasticsearch برای ایندکس‌گذاری و بازیابی سریع داده‌ها و Kibana برای مصورسازی و پایش تحلیل شد. خروجی این فاز، تعیین ابزارها و فناوری‌های مناسب و تدوین معماری اولیه‌ی سیستم بود. البته لازم به ذکر است مطالعه و تحقیق تا پایان این پروژه جزو جدایی ناپذیر آن بود.

فاز دوم: آماده‌سازی مجموعه‌داده BIDS و استخراج دادگان و فراداده‌ها

در این فاز، تمرکز اصلی بر آماده‌سازی و تحلیل اولیه داده‌ها قرار گرفت. ساختار مجموعه‌داده‌های BIDS به صورت دقیق بررسی شد و قابلیت‌های کتابخانه‌ی PyBIDS برای پیمایش و دسترسی به داده‌ها و فراداده‌ها ارزیابی گردید. در نهایت، با بهره‌گیری از PyBIDS و مجموعه‌ای از کتابخانه‌های پایتون، داده‌ها و فراداده‌های مورد نیاز استخراج شده و برای مراحل بعدی پردازش و ایندکس‌گذاری آماده شدند.

فاز سوم: طراحی کدهای ایندکس‌گذاری و ایجاد ایندکس‌های Elasticsearch

در این مرحله، اسکریپت‌های پایتونی توسعه داده شدند تا اطلاعات استخراج‌شده از داده‌ها و فراداده‌ها را به چهار ایندکس مستقل در Elasticsearch منتقل کنند. این کار امکان جستجوی سریع و انعطاف‌پذیر در داده‌ها را فراهم کرد و ساختار اصلی پایگاه داده جستجو محور پروژه را شکل داد.

فاز چهارم: طراحی داشبوردهای Kibana و ارزیابی مصورسازی داده‌ها

در ادامه، سه داشبورد تحلیلی در Kibana طراحی شدند که شامل ایندکس‌های file، subject و info بودند. این داشبوردها ابزار مناسبی برای پایش داده‌ها، نمایش شاخص‌های کلیدی و تحلیل بصری فراهم کردند.

فاز پنجم: توسعه رابط کاربری Flask برای افزودن داده جدید

در این فاز، یک رابط کاربری تحت وب پیاده‌سازی شد که از طریق آن کاربر (پژوهشگر) می‌تواند فایل‌های فشرده حاوی داده‌های آزمودنی‌های جدید را بارگذاری کنند. این داده‌ها به‌طور خودکار در ساختار BIDS ذخیره شده و فراداده آن‌ها به فایل TSV افزوده می‌شود.

فاز ششم: خودکارسازی فرآیندها با Airflow و یکپارچه‌سازی در Docker

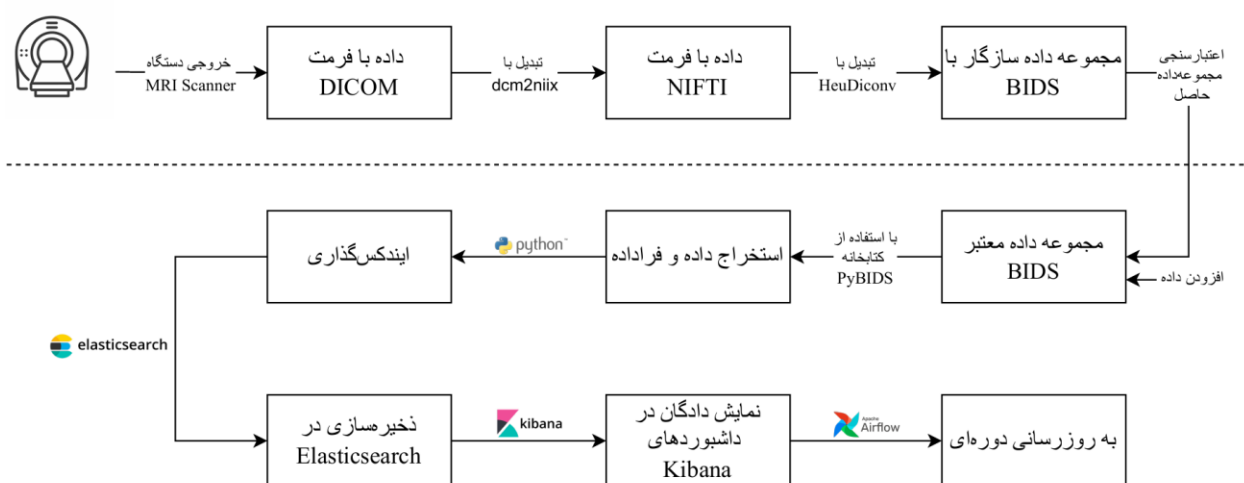
در آخرین فاز، برای هر اسکریپت یک گردش کار در Airflow تعریف شد تا اجرای خودکار وظایف تضمین شود. همچنین کل سامانه در قالب Docker یکپارچه‌سازی شد تا قابلیت استقرار آسان، مدیریت بهتر وابستگی‌ها و مقیاس‌پذیری در محیط‌های مختلف فراهم گردد.

۳-۲-۴ جریان داده‌ها در سیستم

جریان داده‌ها در سیستم پیشنهادی از لحظه ورود داده‌های جدید تا پایش آن‌ها در داشبورد به صورت زیر است:

۱. ورود داده در قالب BIDS به سیستم
۲. استخراج اطلاعات ساختاری و فراداده از BIDS
۳. ایندکس‌گذاری
۴. ذخیره‌سازی در Elasticsearch
۵. نمایش داده در داشبوردهای طراحی شده در Kibana
۶. اجرای گردش کار Airflow برای به‌روزرسانی خودکار
۷. امکان افزودن داده جدید از طریق رابط کاربری Flask

در نمودار شکل ۱۰ جریان داده‌ها نمایش داده شده است که فرایند پس از خط چین در اسکوپ این پروژه جای می‌گیرد.



شکل ۱۰- جریان دادگان از تصویربرداری تا مصورسازی

۳-۳ پیاده‌سازی روش پیشنهادی

۱-۳-۳ استخراج و پردازش داده و فراداده

در این بخش، هدف اصلی استخراج نظام‌مند داده‌ها و فراداده‌ها از مجموعه‌های BIDS و آماده‌سازی آن‌ها برای ایندکس‌گذاری بوده است. ابتدا ساختار استاندارد BIDS به‌طور دقیق مطالعه شد و کتابخانه PyBIDS به‌عنوان ابزار اصلی جهت پیمایش و استخراج اطلاعات انتخاب گردید. همچنین به‌منظور دستیابی به مجموعه داده‌های معتبر، از پایگاه OpenNeuro استفاده شد که دسترسی به مجموعه‌داده‌های متنوع در قالب BIDS را فراهم می‌کند [۵].

فرایند استخراج در سه اسکریپت پایتونی پیاده‌سازی شد:

- در اسکریپت نخست، داده‌های دموگرافیک مربوط به آزمودنی‌ها (از قبیل سن، جنسیت، قد و وزن) از فایل *participants.tsv* استخراج شده و در کنار اطلاعات مربوط به مدالیت‌های هر آزمودنی آماده گردید.

- در اسکریپت دوم، تمرکز بر سطح فایل‌ها بود و فراداده‌ای همچون مسیر فایل، اندازه‌ی فایل و نوع مدالیت‌ها برای کلیه‌ی نمونه‌ها استخراج شد. بخشی از کد این بخش در شکل ۱۱ قابل مشاهده است.

- اسکریپت سوم به پردازش محتوای داخلی فایل‌ها اختصاص یافت؛ به گونه‌ای که برای مدالیت‌های *anat*، *func* و *dwi*، فراداده‌های مرتبط از فایل‌های JSON (مانند قدرت میدان مغناطیسی، زمان تکرار، زمان اکو، زاویه چرخش، نوع تصویر و اطلاعات دستگاه MRI) استخراج گردید. برای مدالیت‌های *beh* نیز اطلاعات رفتاری موجود در فایل‌های TSV مورد پردازش قرار گرفت.

این سه لایه پردازشی مکمل یکدیگر بوده و داده‌های استخراج‌شده را به شکل ساختاریافته و آماده برای انتقال به پایگاه داده‌ی جستجو محور فراهم کردند.

```
# Load layout
layout = BIDSLayout(r"datasets\ds006040", validate=False)

# Prepare bulk data
actions = []
index_name = "file_6040"

files = layout.get()
df = layout.to_df()

for _, row in df.iterrows():
    doc = row.dropna().to_dict()
    file_path = row['path']
    try:
        file_size = os.path.getsize(file_path)
        doc['filesize'] = file_size
    except OSError:
        doc['filesize'] = None

    action = {
        "_index": index_name,
        "_source": doc
    }
    actions.append(action)
```

شکل ۱۱- بخشی از فرایند استخراج فراداده فایل‌ها

۳-۳-۲ ایندکس گذاری و ذخیره سازی

در گام بعد، داده ها و فراداده های استخراج شده در قالب ایندکس های مستقل در Elasticsearch ذخیره سازی شدند. برای این منظور ابتدا سرویس Elasticsearch در محیط Docker مستقر گردید که انعطاف پذیری و سهولت مدیریت زیرساخت را تضمین می کرد. پس از بررسی و تست اتصال از طریق ابزار Postman، عملیات ایندکس گذاری از درون اسکریپت های پایتونی اجرا شد (شکل ۱۲).

هر یک از اسکریپت های سه گانه داده های آماده شده را به ایندکس متناظر خود منتقل کردند: ایندکس آزمودنی ها (subject)، ایندکس فایل ها (file)، ایندکس فراداده های مربوط به مدالیت های تصویری (anat, func, dwi) و ایندکس مربوط به داده های رفتاری (beh). بدین ترتیب ساختاری یکپارچه و جستجو محور شکل گرفت که امکان واکنشی سریع، انعطاف پذیر و هدفمند داده ها را در مراحل بعدی تحلیل و مصورسازی فراهم می کرد.

```
es = Elasticsearch(  
    hosts=[  
        {'host': 'localhost',  
         'port': 9200,  
         'scheme': 'http'  
        }],  
    basic_auth=('elastic', 'changeme')  
)  
  
# Create index  
print(f"index {index_name} created.")  
  
# Bulk insert into Elasticsearch  
bulk(es, actions)  
print(f"all ({len(actions)}) data added to index {index_name}.")
```

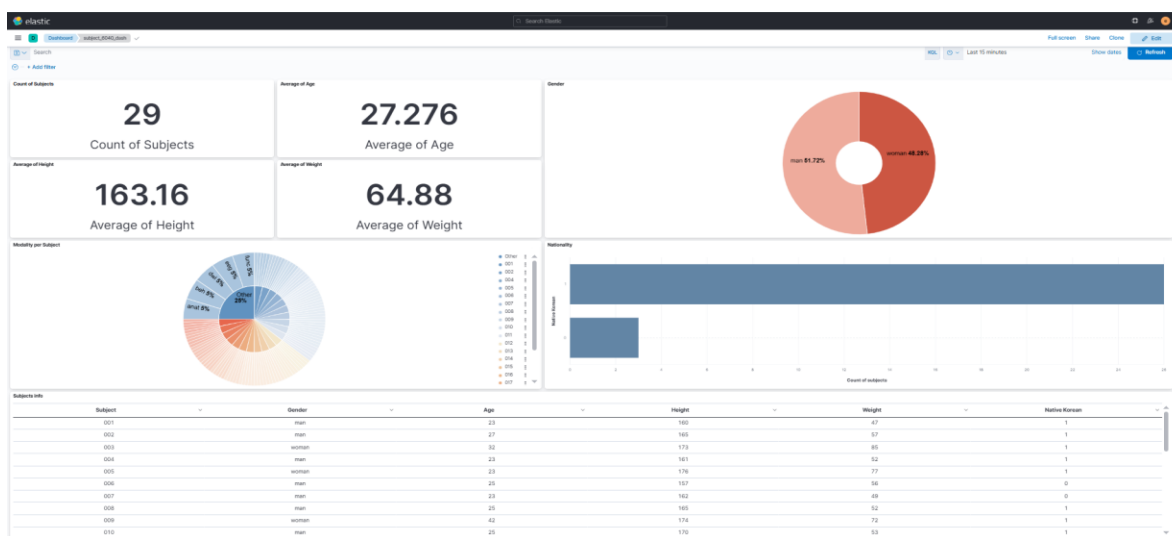
شکل ۱۲- انتقال دادگان استخراج شده به Elasticsearch

۳-۳-۳ پایش و مصورسازی داده‌ها

به منظور فراهم‌سازی امکان تحلیل و پایش داده‌ها، سرویس Kibana در محیط داکر مستقر گردید و اتصال آن به Elasticsearch برقرار شد. این مرحله در ابتدا با چالش‌هایی همراه بود، اما پس از پیکربندی موفق، امکان دسترسی به ایندکس‌های موجود فراهم گردید. برای استفاده از قابلیت‌های مصورسازی Kibana، هر ایندکس به‌صورت یک الگوی شاخص^{۴۵} تعریف شد تا بتواند به‌عنوان منبع داده برای نمودارها و شاخص‌ها مورد استفاده قرار گیرد.

در ادامه، با مطالعه قابلیت‌های مصورسازی و انواع نمودارهای Kibana، سه داشبورد مجزا طراحی و پیاده‌سازی گردید. داشبورد نخست به اطلاعات آزمودنی‌ها اختصاص یافت و شاخص‌های دموگرافیک و کلان داده‌ها را نمایش داد. داشبورد دوم به سطح فایل‌ها پرداخته و توزیع حجم و انواع مدالیت‌ها را مصورسازی کرد. داشبورد سوم نیز بر محتوای فایل‌ها متمرکز بوده و شاخص‌های کلیدی استخراج‌شده از فراداده‌های تصویری (مانند زمان تکرار، زمان اکو، ضخامت برش و نوع تصویر) را به‌صورت ترکیبی از نمودارهای خطی، میله‌ای، دایره‌ای و مقادیر عددی^{۴۶} ارائه داد.

این سه داشبورد در کنار یکدیگر بستری جامع برای پایش داده‌ها و تحلیل بصری فراهم کردند که هم روندهای کلی را آشکار می‌سازد و هم امکان مقایسه‌ی جزئی‌تر میان آزمودنی‌ها و مدالیت‌ها را مهیا می‌کند. (شکل‌های ۱۳ تا ۱۵)

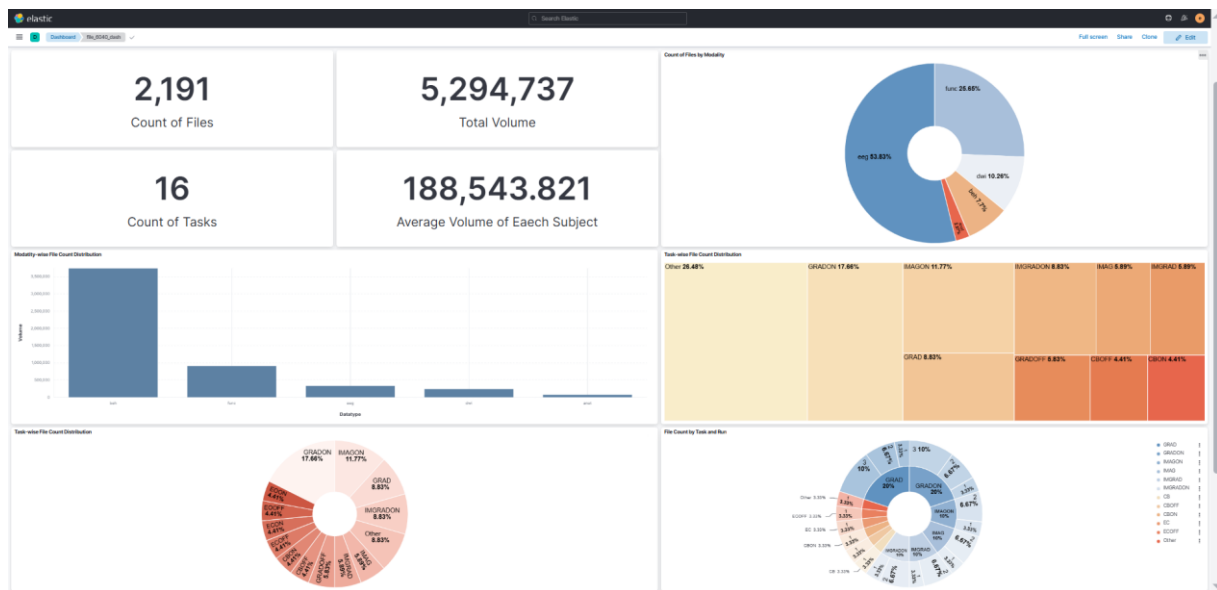


شکل ۱۳- نمای داشبورد نخست (اطلاعات دموگرافیک آزمودنی‌ها)

⁴⁵ Index Pattern

⁴⁶ metric

فصل سوم: روش پیشنهادی و نتیجه گیری



شکل ۱۴- نمای داشبورد دوم (فرا داده فایل‌ها)

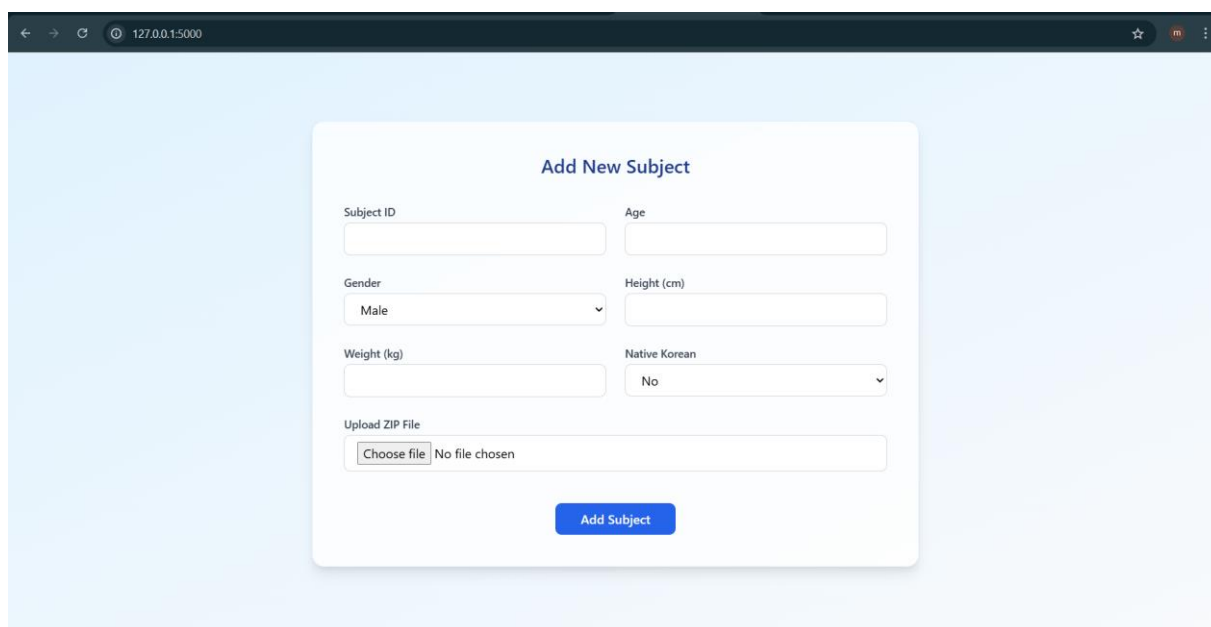


شکل ۱۵- نمای داشبورد سوم (محتوای فایل‌ها)

۳-۳-۴ رابط کاربری برای افزودن داده‌های جدید

برای پویایی بیشتر سیستم و امکان توسعه مداوم پایگاه داده، یک رابط کاربری تحت وب مبتنی بر Flask طراحی شد. این رابط کاربری قابلیت دریافت اطلاعات جمعیت‌شناختی (نظیر شناسه آزمودنی، سن، جنسیت، قد، وزن و متغیرهای تکمیلی) و همچنین بارگذاری فایل فشرده‌ی مربوط به هر آزمودنی را فراهم می‌کند. پس از ارسال داده‌ها، فایل زیپ استخراج شده و در ساختار پوشه‌ی BIDS ذخیره می‌گردد. در ادامه، فراداده واردشده به‌همراه مشخصات آزمودنی در فایل مربوطه اضافه می‌شود تا با سایر اجزای سیستم همگام بماند.

با استفاده از این طراحی، افزودن آزمودنی‌های جدید بدون نیاز به تغییر در ساختار اصلی پروژه امکان‌پذیر شده است. این موضوع سبب می‌شود تا داده‌ها و فراداده‌ها در همان قالب استاندارد BIDS نگهداری شوند و بتوان از کتابخانه‌ها و ابزارهای به‌کاررفته در مراحل پیشین (نظیر PyBIDS و اسکریپت‌های ایندکس‌گذاری) برای پردازش داده‌های جدید نیز استفاده کرد. در شکل ۱۶ نمونه‌ای از رابط کاربری طراحی‌شده برای ثبت آزمودنی‌های جدید نشان داده شده است.



شکل ۱۶- رابط کاربری برای افزودن داده

۳-۳-۵ مازول به روزرسانی خودکار با Airflow

به منظور تضمین به روزرسانی منظم داده‌ها در Elasticsearch و نمایش تغییرات در داشبوردهای Kibana، از ابزار Apache Airflow استفاده شد. همان‌گونه که در بخش رابط کاربری بیان شد، داده‌های جدید تنها به مجموعه داده اضافه می‌شوند و مستقیماً در ایندکس‌های Elasticsearch قرار نمی‌گیرند. بنابراین، لازم است فرآیند ایندکس‌گذاری به صورت دوره‌ای و خودکار اجرا شود.

برای این منظور، سه اسکریپت استخراج داده‌ها که پیش‌تر طراحی شده بودند، به شکل سه گردش کار مستقل در Airflow پیاده‌سازی شدند. هر گردش کار وظیفه‌ی اجرای روزانه‌ی یک بخش از فرآیند ایندکس‌گذاری (آزمودنی‌ها، فایل‌ها و فراداده) را بر عهده دارد. با این مکانیزم، هرگونه تغییر یا داده‌ی افزوده‌شده در پایگاه داده اصلی به طور خودکار در Elasticsearch نیز بازتاب می‌یابد و نتایج به روزرسانی شده در Kibana قابل مشاهده خواهد بود. محیط رابط کاربری این سرویس در شکل ۱۷ قابل مشاهده است. در این محیط گردش کارها به همراه سابقه تعداد دفعات اجرای موفق یا ناموفق هر یک نمایش داده می‌شوند. همچنین قابلیت تحریک کردن هر گردش کار به شکل دستی در این محیط وجود دارد. برای مثال شکل ۱۷ نشان می‌دهد اولین گردش کار تا کنون پنج اجرای موفق و یک اجرای ناموفق داشته است.

DAG	Owner	Runs	Schedule	Last Run	Next Run	Recent Tasks	Actions	Links
bids_files_6040	airflow	1 (Success) 1 (Failure)	@daily	2025-08-27, 00:00:00	2025-08-28, 00:00:00	1 (Success) 1 (Failure)	[Play] [Stop] [Refresh]	...
bids_info_6040	airflow	1 (Success)	None	2025-08-12, 16:03:21		1 (Success)	[Play] [Stop] [Refresh]	...
bids_subjects_6040	airflow	20 (Success) 1 (Failure)	@daily	2025-08-27, 00:00:00	2025-08-28, 00:00:00	1 (Success)	[Play] [Stop] [Refresh]	...

شکل ۱۷- رابط کاربری سرویس Apache Airflow

نمونه‌ای از تعریف یک گردش کار در Airflow برای بارگذاری اطلاعات فایل‌های مجموعه داده در Elasticsearch در کد شکل ۱۸ ارائه شده است.

```
with DAG(  
    dag_id="bids_files_6040",  
    default_args=default_args,  
    schedule_interval='@daily',  
    catchup=False,  
    description="Upload dataset files metadata to Elasticsearch",  
) as dag:  
  
    upload_task = PythonOperator(  
        task_id="load_bids_files_to_es",  
        python_callable=load_files_to_elasticsearch,  
    )  
  
    upload_task
```

شکل ۱۸ - نمونه‌ای از تعریف یک گردش کار

این رویکرد موجب می‌شود که همگام‌سازی میان مجموعه داده، Elasticsearch و داشبوردهای Kibana به صورت خودکار، شفاف و پایدار انجام گیرد.

۳-۳-۶ یکپارچه‌سازی ماژول‌ها در زیرساخت Docker

به منظور استقرار یکپارچه و پایدار اجزای مختلف سیستم، از Docker-Compose برای مدیریت سرویس‌ها استفاده شد. در گام نخست، سرویس‌های Airflow و اجزای مرتبط با آن (وب سرور، Scheduler و Worker) به صورت جداگانه از Elasticsearch و Kibana اجرا شدند. اما این تفکیک منجر به بروز مشکلات ارتباطی در سطح شبکه شد و سرویس Airflow قادر به برقراری اتصال مستقیم با سرویس Elasticsearch نبود.

برای رفع این مشکل، کلیه سرویس‌ها در قالب یک پیکربندی یکپارچه‌ی Docker-Compose و در بستر یک شبکه‌ی داخلی مشترک استقرار یافتند تا ارتباطات میان آن‌ها به صورت پایدار و قابل اطمینان برقرار گردد. این معماری امکان برقراری ارتباط پایدار میان Airflow، پایگاه داده PostgreSQL و نیز Elasticsearch و Kibana را فراهم

ساخت. در این ساختار، Airflow وظیفه‌ی زمان‌بندی و اجرای وظایف ایندکس‌گذاری را برعهده دارد، در حالی که Elasticsearch به عنوان موتور جستجو و Kibana به عنوان ابزار مصورسازی داده‌ها عمل می‌کنند. این یکپارچگی، علاوه بر رفع مشکلات شبکه‌ای، موجب سهولت در استقرار، مقیاس‌پذیری و مدیریت یکپارچه اجزای سیستم گردید.

۳-۴ روش ارزیابی

۳-۴-۱ نوع داده‌ها و مجموعه‌داده مورد استفاده برای ارزیابی

برای ارزیابی روش پیشنهادی، از مجموعه‌داده *OpenNeuro – Sustained Attention Task (gradCPT)* با شماره دسترسی *ds006040* استفاده شد [۱۷]. این مجموعه‌داده شامل داده‌های fMRI، EEG و DWI از ۲۸ شرکت‌کننده در بازه سنی ۱۹ تا ۴۲ سال است. داده‌های fMRI و DWI با دستگاه ۳ T Siemens Magnetom Prisma جمع‌آوری شده و EEG با استفاده از ۶۴ کانال Brain Product BrainCap MR ثبت شده است.

مجموعه شامل یک جلسه^{۴۷} برای هر شرکت‌کننده و طیف متنوعی از تسک‌ها است؛ از جمله وظیفه‌ی حالت استراحت^{۴۸} (با چشم باز و بسته)، الگوی شطرنجی با فرکانس ۱۵ هرتز، وظیفه‌ی تداعی تدریجی^{۴۹} و وظیفه‌ی تجسم ذهنی^{۵۰}. وجود این تنوع در تسک‌ها و مدالیت‌ها، امکان تولید و تحلیل داشبوردهای متنوع در Kibana را فراهم می‌کند.

انتخاب این مجموعه‌داده به چند دلیل صورت گرفت:

- حجم مناسب داده و تعداد قابل قبول آزمودنی‌ها، که شرایط واقعی‌تری برای تست زیرساخت فراهم می‌کرد.
- وجود چندین مدالیت و تسک متنوع در مقایسه با بسیاری از مجموعه‌داده‌های مشابه، که امکان محک زدن سیستم در سناریوهای مختلف و طراحی داشبوردهای چندبعدی را فراهم می‌ساخت.
- ساختار استاندارد BIDS، که موجب شد فرآیند یکپارچه‌سازی داده‌ها با Airflow و Elasticsearch به صورت ساده‌تر و سازگارتر انجام گیرد.

⁴⁷ Session

⁴⁸ Resting-state

⁴⁹ gradCPT

⁵⁰ Imagery task

این ویژگی‌ها باعث شد مجموعه داده انتخابی مبنای مناسبی برای ارزیابی قابلیت‌های سیستم طراحی شده، شامل به‌روزرسانی خودکار، شاخص‌های جستجو و تحلیل، و نمایش داشبوردهای پژوهشی قرار گیرد.

۳-۴-۲ سناریوهای ارزیابی عملکرد

به منظور سنجش کارایی زیرساخت طراحی شده، سه سناریوی اصلی تعریف گردید که هر کدام بخشی از چرخه‌ی پردازش، ایندکس‌گذاری و تحلیل داده‌ها را مورد آزمون قرار می‌دهند. این سناریوها به گونه‌ای انتخاب شده‌اند که از یک سو عملکرد فنی اجزای کلیدی سیستم (شامل Elasticsearch، Airflow و Kibana) را پوشش دهند و از سوی دیگر، بازتابی از نیازهای واقعی پژوهشگران در کار با داده‌های علوم اعصاب باشند.

معیارهای مورد سنجش در این سناریوها عبارت‌اند از:

- پویایی سیستم در مواجهه با ورود داده‌های جدید و انعکاس به‌موقع آن‌ها در ایندکس‌ها و داشبوردها
- کارایی در جستجو و بازیابی اطلاعات از نظر سرعت و دقت در پاسخ به پرس‌وجوها
- پایداری و صحت به‌روزرسانی‌های دوره‌ای در فرآیند اجرای خودکار گردش کارها و حفظ یکپارچگی داده‌ها

در ادامه، هر سناریو به صورت جداگانه معرفی و تشریح می‌شود.

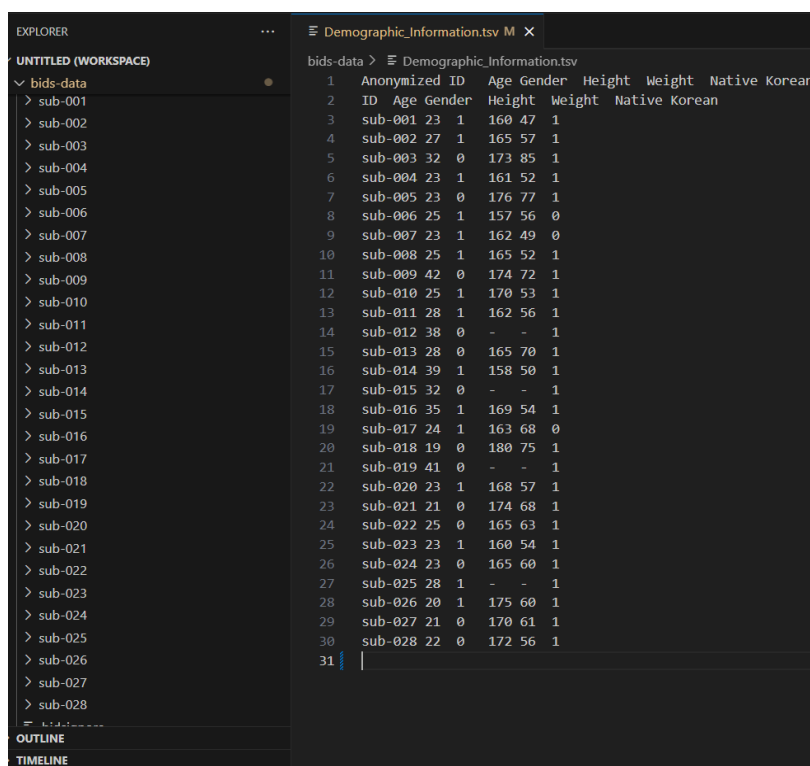
۳-۴-۲-۱ سناریو اول، افزودن داده جدید

برای سنجش پویایی سیستم، یک چرخه‌ی کامل از وضعیت اولیه، افزودن داده جدید سپس به‌روزرسانی خودکار و نهایتاً مشاهده تغییرات در داشبوردها پیاده‌سازی می‌شود. معیار موفقیت این سناریو آن است که افزایش تعداد آزمودنی‌ها در سطح مجموعه داده، پس از اجرای گردش کار در Airflow، به‌طور کامل در شاخص‌های ذخیره‌شده در Elasticsearch و نمایش در Kibana بازتاب یابد.

۱. مرحله اول: نمایش حالت اولیه

ابتدا تعداد آزمودنی‌ها را در سه سطح بررسی می‌کنیم:

- سطح اول) مجموعه داده خام (ساختار پوشه‌ها): همانطور که در شکل ۱۹ مشاهده می‌شود، تعداد آزمودنی‌ها هم در فایل TSV و هم در میان پوشه‌های هر آزمودنی، ۲۸ می‌باشد.

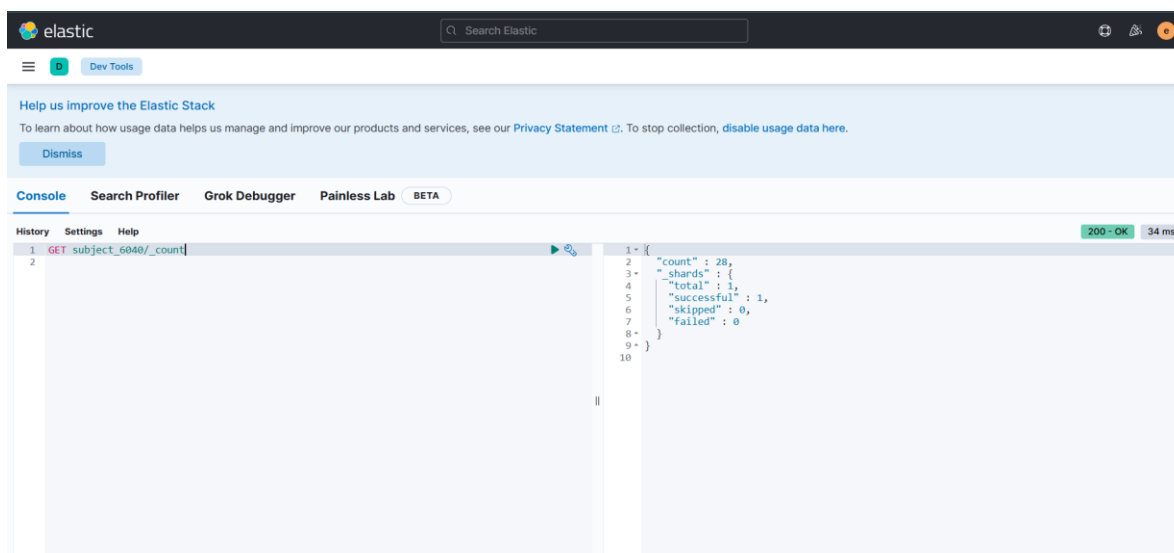


	Anonymized ID	Age	Gender	Height	Weight	Native Korean
1	ID	Age	Gender	Height	Weight	Native Korean
2	sub-001	23	1	160	47	1
3	sub-002	27	1	165	57	1
4	sub-003	32	0	173	85	1
5	sub-004	23	1	161	52	1
6	sub-005	23	0	176	77	1
7	sub-006	25	1	157	56	0
8	sub-007	23	1	162	49	0
9	sub-008	25	1	165	52	1
10	sub-009	42	0	174	72	1
11	sub-010	25	1	170	53	1
12	sub-011	28	1	162	56	1
13	sub-012	38	0	-	-	1
14	sub-013	28	0	165	70	1
15	sub-014	39	1	158	50	1
16	sub-015	32	0	-	-	1
17	sub-016	35	1	169	54	1
18	sub-017	24	1	163	68	0
19	sub-018	19	0	180	75	1
20	sub-019	41	0	-	-	1
21	sub-020	23	1	168	57	1
22	sub-021	21	0	174	68	1
23	sub-022	25	0	165	63	1
24	sub-023	23	1	160	54	1
25	sub-024	23	0	165	60	1
26	sub-025	28	1	-	-	1
27	sub-026	20	1	175	60	1
28	sub-027	21	0	170	61	1
29	sub-028	22	0	172	56	1
30						
31						

شکل ۱۹- مجموعه داده خام پیش از افزودن داده

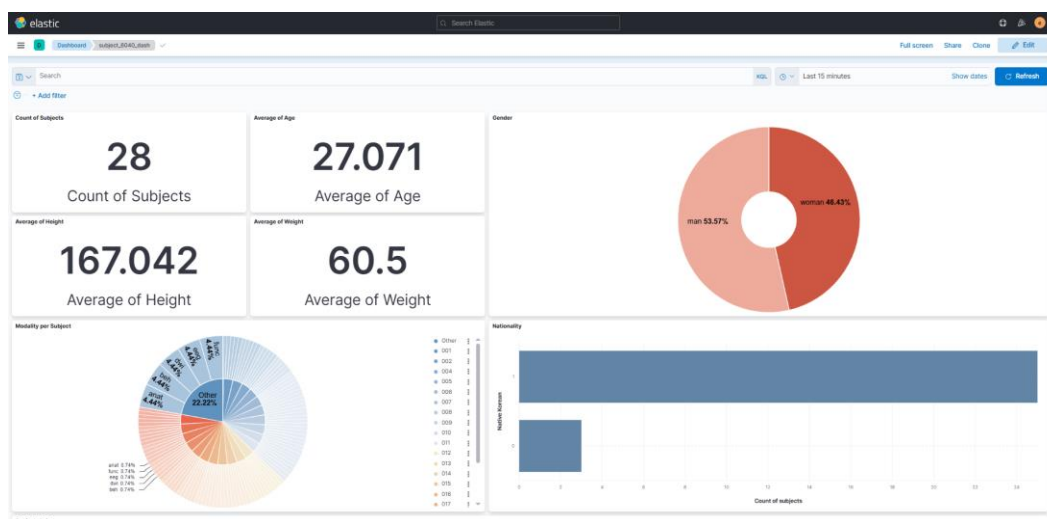
فصل سوم: روش پیشنهادی و نتیجه گیری

- سطح دوم) ایندکس‌های Elasticsearch: همانطور که در شکل ۲۰ مشاهده می‌شود، تعداد آزمودنی‌ها در ایندکس اختصاص یافته به آنها، ۲۸ می‌باشد.



شکل ۲۰- پرس‌وجو برای شمارش آزمودنی‌ها در ایندکس subject_6040

- سطح سوم) داشبورد Kibana: در این سطح نیز مشاهده می‌شود تعداد آزمودنی‌ها ۲۸، گزارش شده است.

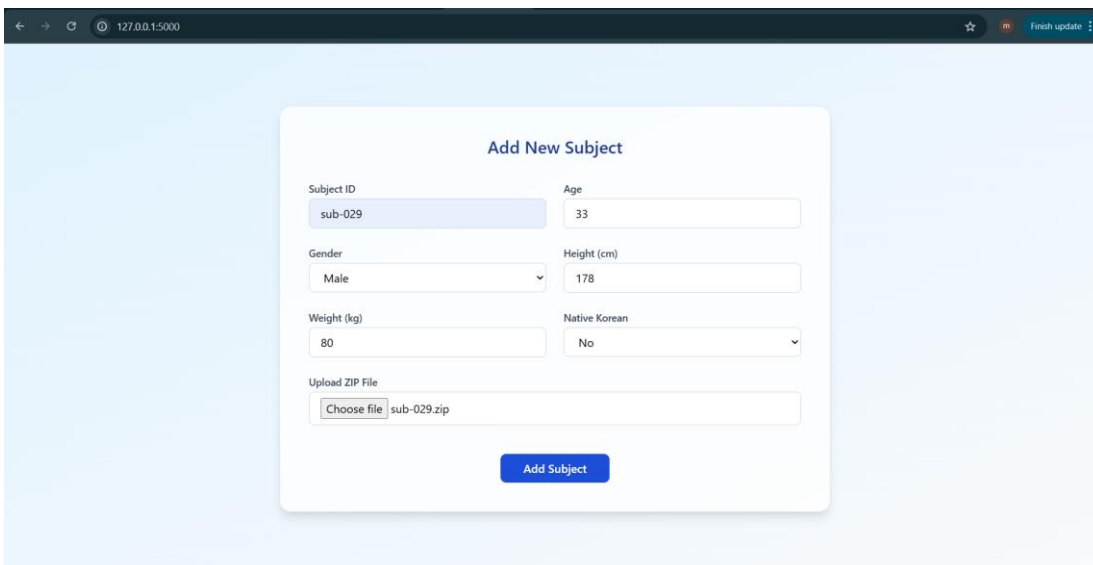


شکل ۲۱- نمای داشبورد آزمودنی‌ها

فصل سوم: روش پیشنهادی و نتیجه گیری

۲. مرحله دوم: افزودن داده از طریق رابط کاربری

در این مرحله از ارزیابی، مطابق شکل ۲۲ مشخصات آزمودنی جدید به همراه فایل فشرده دادگان تصویربرداری آزمودنی را در فرم رابط کاربری طراحی شده ارسال می‌کنیم.

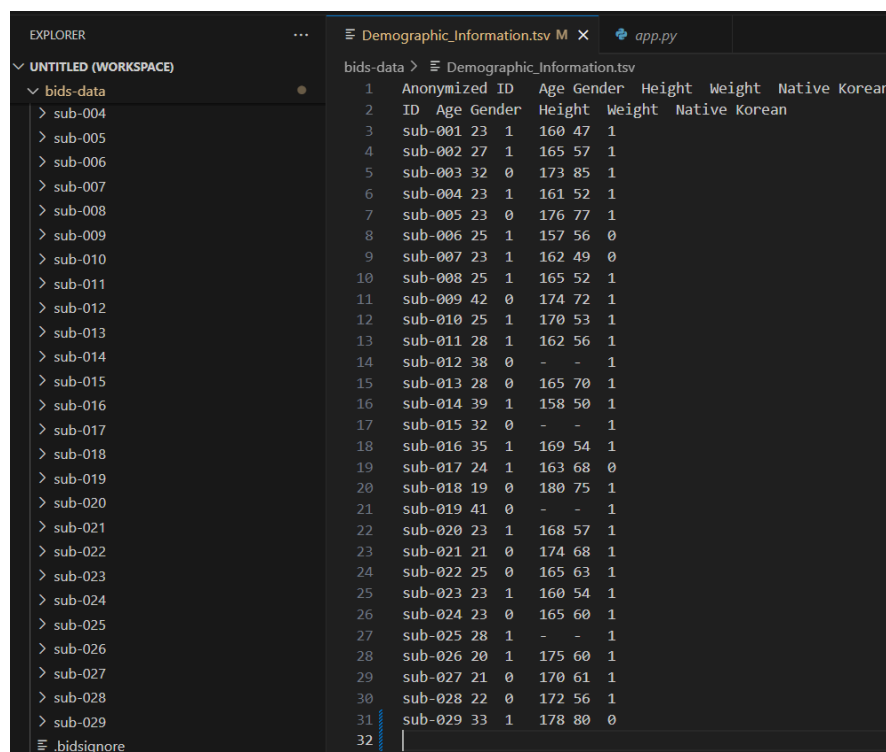


The screenshot shows a web browser window with the address bar displaying '127.0.0.1:5000'. The main content area features a light blue background with a white modal form titled 'Add New Subject'. The form has several input fields: 'Subject ID' with the value 'sub-029', 'Age' with '33', 'Gender' with a dropdown menu showing 'Male', 'Height (cm)' with '178', 'Weight (kg)' with '80', and 'Native Korean' with a dropdown menu showing 'No'. Below these is a file upload section labeled 'Upload ZIP File' with a 'Choose file' button and the filename 'sub-029.zip'. At the bottom of the form is a blue button labeled 'Add Subject'.

شکل ۲۲ - افزودن داده از طریق رابط کاربری

۳. مرحله سوم: بررسی مجدد مجموعه داده

ساختار مجموعه داده مجدداً بررسی شده و مشاهده می‌شود که تعداد آزمودنی‌ها به ۲۹ رسیده است که در شکل ۲۳ قابل مشاهده است.



The screenshot shows a code editor with a file named 'Demographic_Information.tsv' open. The file contains demographic information for 32 subjects, organized into a table with columns: Anonymized ID, Age, Gender, Height, Weight, and Native Korean. The subjects are listed from 1 to 32, with some having missing values for Height and Weight (indicated by dashes). The file is saved as 'app.py'.

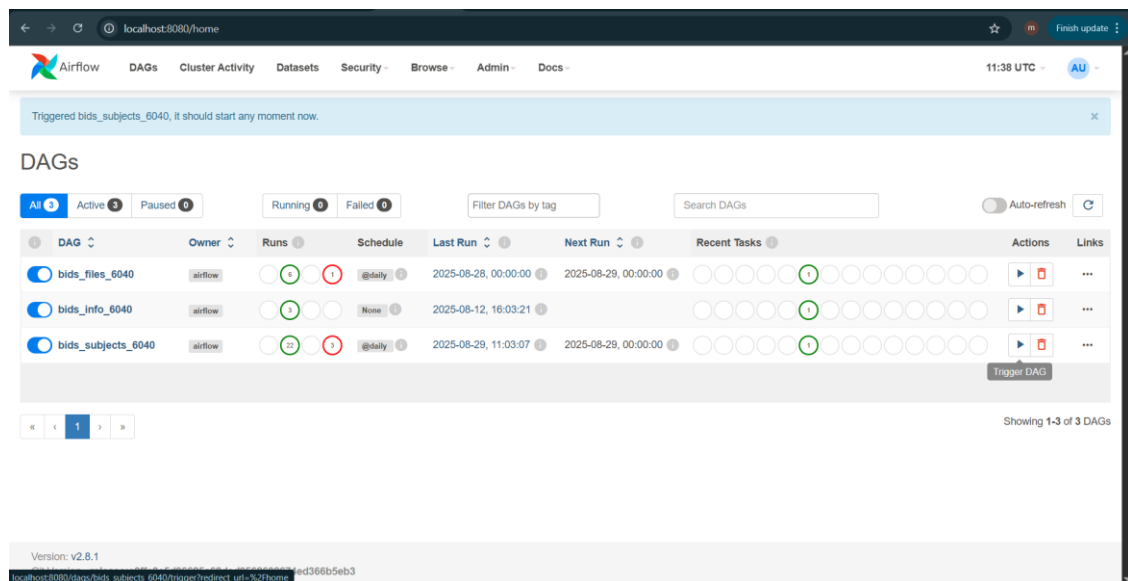
Anonymized ID	Age	Gender	Height	Weight	Native Korean
sub-001	23	1	160	47	1
sub-002	27	1	165	57	1
sub-003	32	0	173	85	1
sub-004	23	1	161	52	1
sub-005	23	0	176	77	1
sub-006	25	1	157	56	0
sub-007	23	1	162	49	0
sub-008	25	1	165	52	1
sub-009	42	0	174	72	1
sub-010	25	1	170	53	1
sub-011	28	1	162	56	1
sub-012	38	0	-	-	1
sub-013	28	0	165	70	1
sub-014	39	1	158	50	1
sub-015	32	0	-	-	1
sub-016	35	1	169	54	1
sub-017	24	1	163	68	0
sub-018	19	0	180	75	1
sub-019	41	0	-	-	1
sub-020	23	1	168	57	1
sub-021	21	0	174	68	1
sub-022	25	0	165	63	1
sub-023	23	1	160	54	1
sub-024	23	0	165	60	1
sub-025	28	1	-	-	1
sub-026	20	1	175	60	1
sub-027	21	0	170	61	1
sub-028	22	0	172	56	1
sub-029	33	1	178	80	0

شکل ۲۳- مجموعه داده خام پس از افزودن داده

۴. مرحله چهارم: اجرای گردش کار

گردش کار مربوطه مطابق شکل ۲۴ در Airflow فعال سازی می شود تا ایندکس ها آپدیت و داده ها وارد Elasticsearch شوند.

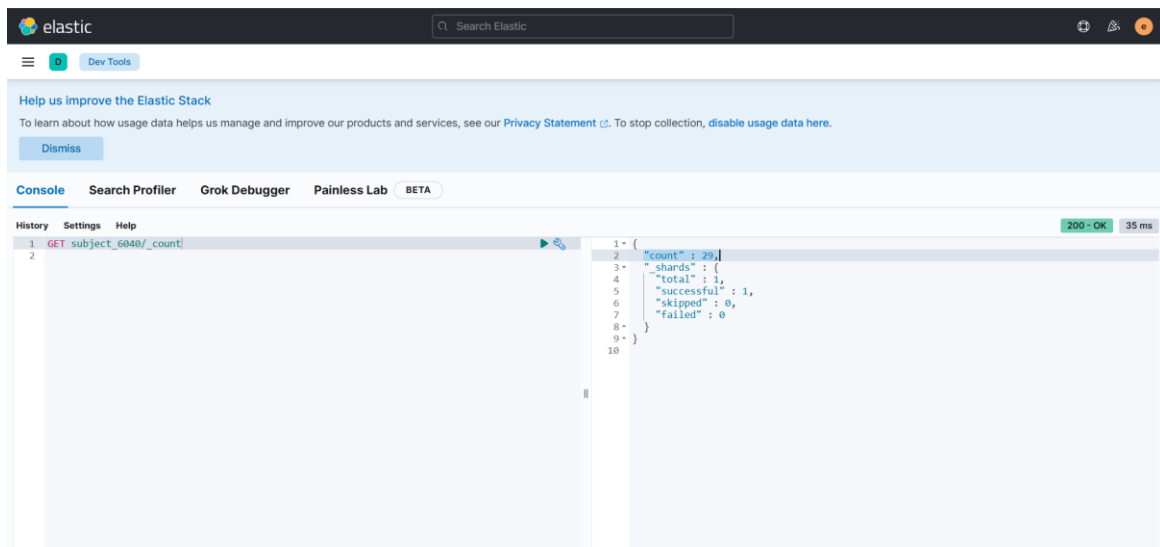
فصل سوم: روش پیشنهادی و نتیجه گیری



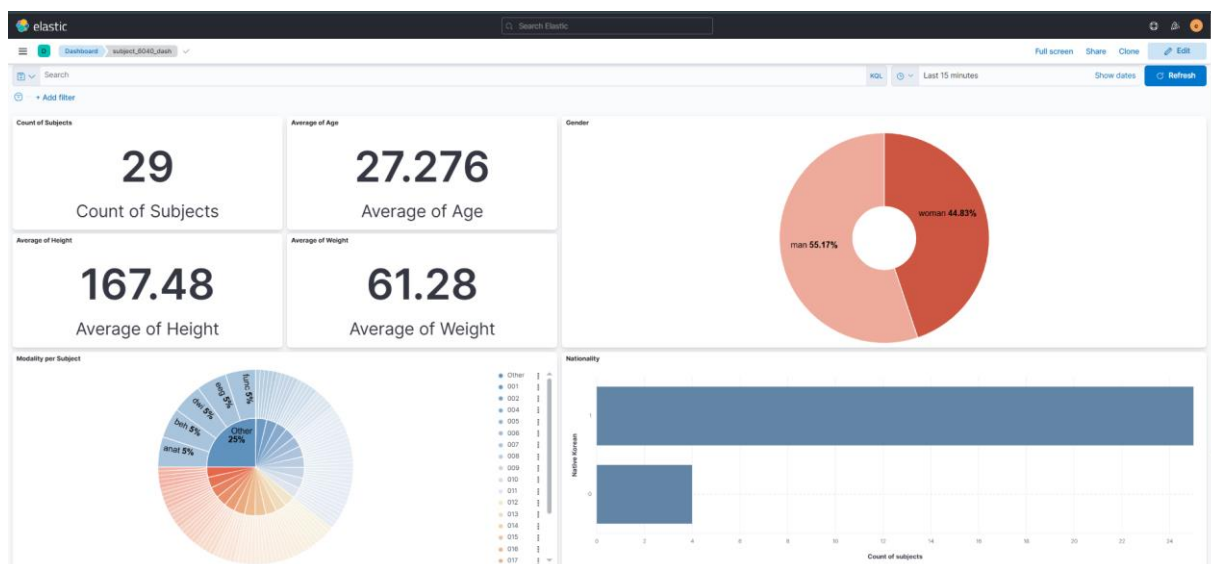
شکل ۲۴- اجرای گردش کار آزمودنی‌ها در Airflow

۵. مرحله پنجم: نمایش وضعیت به‌روزشده در Elasticsearch و Kibana

با اجرای یک پرس‌وجو ساده در Elasticsearch تعداد آزمودنی‌ها شمارش شده و مقدار جدید (۲۹) به نمایش درمی‌آید (شکل ۲۵). همچنین در Kibana داشبورد به روزشده نشان داده می‌شود (شکل ۲۶).



شکل ۲۵- پرس‌وجو برای شمارش آزمودنی‌ها در ایندکس subject_6040



شکل ۲۶- نمای داشبورد آزمودنی‌ها

۳-۲-۴-۲ سناریو دوم، جستجو و بازیابی اطلاعات

در این سناریو هدف آن است که کارایی سیستم پیشنهادی در انجام عملیات جستجو و بازیابی اطلاعات ارزیابی شود. از آن جا که یکی از مهم‌ترین نیازمندی‌های یک بایوبانک داده‌های تصویربرداری مغزی، امکان دسترسی سریع و دقیق به فایل‌ها و فراداده‌های مرتبط است، مقایسه‌ی دو رویکرد متفاوت در بازیابی داده‌ها اهمیت دارد. به همین منظور، یک پرس‌وجو مشخص همزمان با استفاده از PyBIDS و Elasticsearch اجرا می‌شود تا تفاوت زمان پاسخ و صحت نتایج بررسی گردد. این مقایسه نشان می‌دهد که اگرچه PyBIDS برای مدیریت ساختار استاندارد BIDS طراحی شده است، اما Elasticsearch به دلیل ایندکس‌گذاری پیشرفته و موتور جستجوی بهینه، سرعت و انعطاف‌پذیری بسیار بیشتری در پاسخ به پرس‌وجوها ارائه می‌دهد.

۱. جستجوی با کتابخانه PyBIDS:

با استفاده از کد موجود در شکل ۲۷ از کتابخانه PyBIDS برای جستجوی همه فایل‌های تصویربرداری ساختاری^{۵۱} با پسوند *nii.gz* استفاده کرده‌ایم. PyBIDS کل ساختار فایل‌ها را پیمایش می‌کند، سپس با توجه به فراداده استخراج‌شده نتایج را برمی‌گرداند.

```
import time
from bids import BIDSLayout

start = time.time()

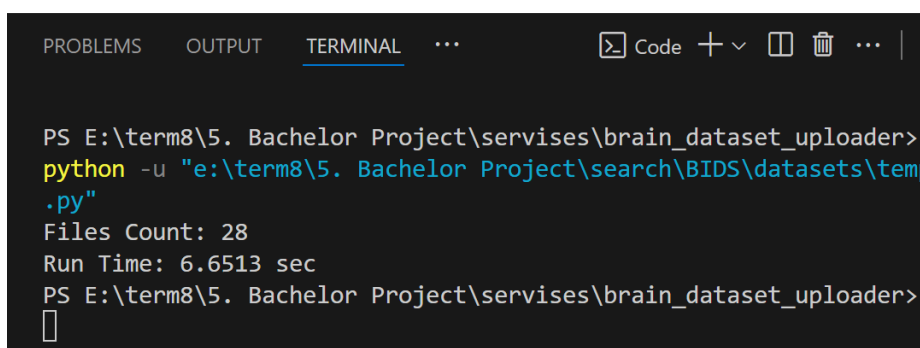
layout = BIDSLayout(r"services\bids-data")
anat_files = layout.get(datatype="anat", extension="nii.gz")

end = time.time()

print(f"Files Count: {len(anat_files)}")
print(f"Run Time: {end - start:.4f} sec")
```

شکل ۲۷ - کد اجرای پرس‌وجو نمونه با استفاده از کتابخانه PyBIDS

نتیجه اجرای این پرس‌وجو در شکل ۲۸ قابل مشاهده است. اجرای این پرس‌وجو ۶.۶۵۱۳ ثانیه زمان برده و نتیجه حاصل، یافتن ۲۸ سند مربوطه است.



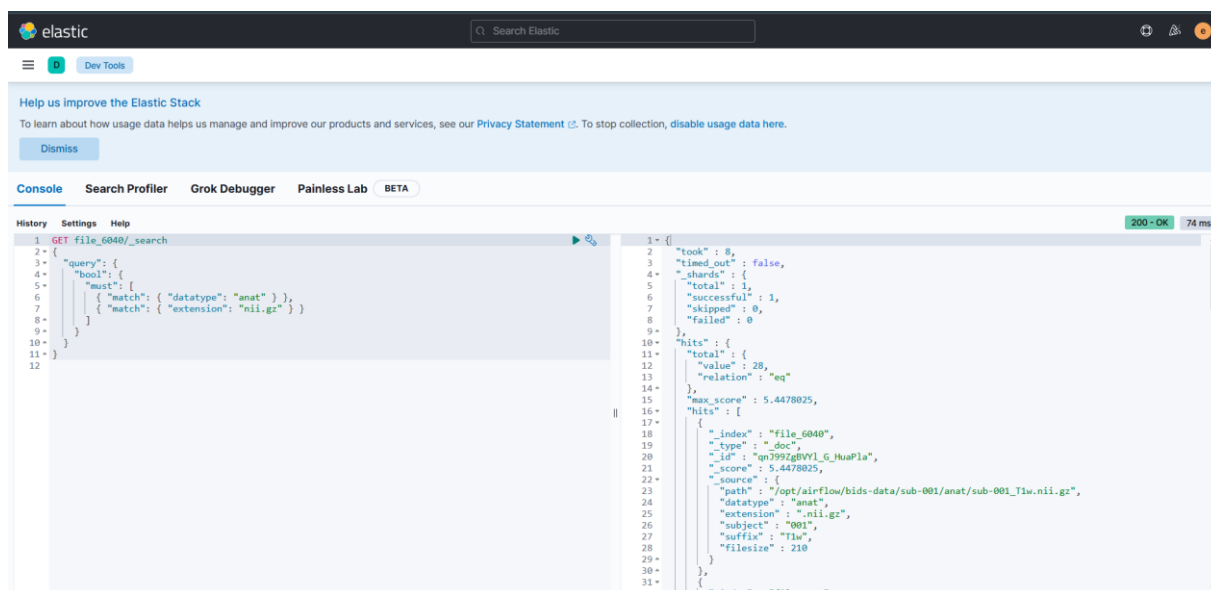
```
PROBLEMS  OUTPUT  TERMINAL  ...  [Code] + - [Icons] |
PS E:\term8\5. Bachelor Project\services\brain_dataset_uploader>
python -u "e:\term8\5. Bachelor Project\search\BIDS\datasets\temp
.py"
Files Count: 28
Run Time: 6.6513 sec
PS E:\term8\5. Bachelor Project\services\brain_dataset_uploader>
█
```

شکل ۲۸ - نتیجه اجرای پرس‌وجو نمونه با استفاده از کتابخانه PyBIDS

⁵¹ anat

۲. جستجو با موتور جستجوی Elasticsearch:

پرسوجو معادل را در محیط Dev tools اجرا می کنیم. همانطور که در شکل ۲۹ می بینیم، پرسوجو در سمت چپ و نتیجه آن در سمت راست شکل قابل مشاهده است. از تفسیر نتیجه می توان دریافت بخش 8 : "took" یعنی اجرای این پرسوجو ۸ میلی ثانیه طول کشیده است و بخش 28 : "value" بیانگر این است که ۲۸ سند در ایندکس file_6040 با شرایط داده شده پیدا شده اند.



شکل ۲۹- پرسوجو برای شمارش آزمودنی ها پس از افزودن داده

نتیجه مقایسه این دو روش:

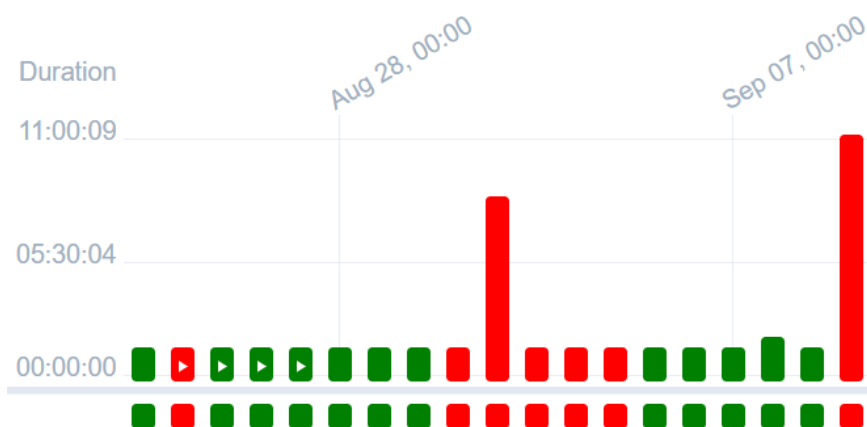
جدول ۱- مقایسه نتایج دو روش جستجو با PyBIDS و Elasticsearch

Elasticsearch	PyBIDS	
۲۸	۲۸	تعداد سند یافت شده
۰/۰۰۸	۶/۶۵۱۳	زمان اجرا (ثانیه)

در این سناریو هر دو روش، یعنی PyBIDS و Elasticsearch، یک وظیفه‌ی یکسان را انجام داده‌اند: جستجوی فایل‌های تصویربرداری ساختاری با پسوند nii.gz و بازگرداندن نتایج منطبق. خروجی هر دو نیز کاملاً مشابه بوده و در هر دو حالت ۲۸ سند مرتبط شناسایی شده است. با این حال، تفاوت اصلی در زمان پاسخ‌گویی نمایان می‌شود. در حالی که اجرای پرس‌وجو در PyBIDS حدود ۶.۶۵ ثانیه به طول انجامیده است، همان پرس‌وجو در Elasticsearch تنها در ۸ میلی‌ثانیه پاسخ داده است که یعنی Elasticsearch حدود ۸۳۰ برابر سریعتر از PyBIDS عمل کرده است. این اختلاف چشمگیر نشان می‌دهد که استفاده از ایندکس‌های بهینه و موتور جستجوی معکوس در Elasticsearch سرعت بازیابی اطلاعات را به طور قابل توجهی افزایش می‌دهد و آن را برای سناریوهای عملی با داده‌های حجیم، بسیار مناسب‌تر از روش‌های مبتنی بر پیمایش مستقیم ساختار فایل‌ها مانند PyBIDS می‌سازد.

۳-۲-۴-۳ سناریو سوم، پایداری و صحت به‌روزرسانی دوره‌ای

در سناریو سوم، هدف بررسی پایداری و صحت به‌روزرسانی‌های دوره‌ای است. برای این منظور، وضعیت اجرای سیستم طی چند روز متوالی بررسی شد و عملکرد گردش کارهای Airflow در ایندکس‌گذاری خودکار داده‌ها پایش گردید. معیار اصلی این سناریو اطمینان از اجرای بدون خطا، به‌روزرسانی صحیح ایندکس‌ها و حفظ یکپارچگی داده‌ها در Elasticsearch است. در ادامه به بررسی پایداری سیستم و نظم اجرای اسکریپت‌ها می‌پردازیم.



شکل ۳۰- روند اجرای دوره‌ای

همانطور که در شکل ۳۰ مشاهده می‌شود، در این آزمایش، طی بازه‌ی زمانی مشخص تعداد ۱۹ اجرای متوالی از گردش کارهای تعریف‌شده در Airflow ثبت شد. از میان این اجراها، ۱۲ اجرای موفق و ۷ مورد ناموفق مشاهده گردید. بررسی علل خطاها نشان داد که عدم موفقیت مربوط به معماری یا طراحی سیستم نبوده، بلکه ناشی از شرایط اجرای

محلی^{۵۲} شامل خاموش بودن دستگاه یا در دسترس نبودن سرویس‌های زیرساختی بر روی داکر در زمان‌بندی‌های مشخص بوده است. بنابراین می‌توان نتیجه گرفت که در صورت استقرار همین معماری بر روی یک بستر پایدار سروری^{۵۳}، انتظار می‌رود تمامی اجرای‌های زمان‌بندی‌شده بدون خطا و با حفظ صحت ایندکس‌گذاری در Elasticsearch انجام پذیرد. این امر نشان می‌دهد که روش پیشنهادی از منظر پایداری^{۵۴} و قابلیت اطمینان^{۵۵} در به‌روزرسانی‌های دوره‌ای، مناسب و قابل اتکا است.

۳-۵ نتیجه‌گیری

ایده اولیه این پژوهش طراحی و پیاده‌سازی یک بستر یکپارچه برای مدیریت، ایندکس‌گذاری و بازیابی داده‌های تصویربرداری مغزی در قالب استاندارد BIDS بود. هدف آن بود که علاوه بر فراهم کردن امکان ذخیره‌سازی منظم داده‌ها، قابلیت‌هایی همچون پویایی در افزودن داده‌های جدید، سرعت بالا در جستجو و بازیابی، و پایداری در به‌روزرسانی‌های دوره‌ای نیز تضمین شود. برای دستیابی به این اهداف، معماری پیشنهادی شامل لایه‌های پردازش داده، ایندکس‌گذاری در Elasticsearch، مدیریت گردش کار با Airflow و مصورسازی در Kibana پیاده‌سازی شد.

نتایج ارزیابی در سه سناریوی تعریف‌شده نشان داد که اهداف پروژه تا حد زیادی محقق شده است. در سناریوی نخست، توانایی سیستم در شناسایی و انعکاس تغییرات ناشی از افزودن داده‌های جدید در تمامی سطوح (مجموعه‌داده خام، ایندکس‌ها و داشبوردها) تأیید شد. این نتیجه نشان می‌دهد که معماری طراحی‌شده از پویایی کافی برخوردار است و می‌تواند رشد تدریجی داده‌ها را مدیریت کند.

در سناریوی دوم، مقایسه‌ی PyBIDS و Elasticsearch در یک وظیفه‌ی جستجوی یکسان، بیانگر برتری چشمگیر Elasticsearch در زمان پاسخ‌گویی (حدود ۸۳۰ برابر سریع‌تر) بود، در حالی که هر دو روش از منظر دقت و صحت نتایج عملکرد یکسانی داشتند. این امر مؤید آن است که استفاده از موتور جستجوی ایندکس‌محور نه تنها برای کار با مجموعه‌داده‌های کوچک بلکه به‌ویژه در مواجهه با داده‌های حجیم، کارایی بالاتری خواهد داشت.

⁵² Local execution

⁵³ Stable server environment

⁵⁴ Stability

⁵⁵ Reliability

در سناریوی سوم، اگرچه بخشی از اجرای‌های دوره‌ای با خطا مواجه شد، اما تحلیل‌ها نشان داد که این مسئله ناشی از محدودیت‌های محیط محلی و نه معماری پیشنهادی بوده است. بنابراین می‌توان نتیجه گرفت که در صورت استقرار سیستم بر روی یک سرور پایدار، مکانیزم به‌روزرسانی دوره‌ای بدون خطا و به‌صورت قابل اتکا عمل خواهد کرد.

جدول ۲، خلاصه‌ای از نتایج ارزیابی را در سه سناریو نشان می‌دهد:

جدول ۲- نتایج ارزیابی

سناریو	هدف ارزیابی	نتیجه ارزیابی	تفسیر
۱ افزودن داده جدید	بررسی پویایی و انعکاس تغییرات	داده‌ی جدید در تمام سطوح شناسایی و به‌روز شد	معماری توانایی مدیریت رشد تدریجی داده‌ها را دارد
۲ جستجو و بازیابی	مقایسه PyBIDS و Elasticsearch	صحت نتایج یکسان؛ سرعت ۸۳۰ برابر بیشتر	مناسب برای کاربردهای عملی و داده‌های حجیم
۳ به‌روزرسانی دوره‌ای	ارزیابی پایداری اجرای خودکار	۱۲ موفق، ۷ ناموفق	روی سرور پایدار انتظار اجرای بدون خطا وجود دارد

به طور کلی می‌توان گفت معماری پیشنهادی، ضمن حفظ سازگاری با استاندارد BIDS، امکان مدیریت کارآمد داده‌ها را فراهم کرده و بهبود چشمگیری در سرعت و انعطاف‌پذیری بازیابی اطلاعات ارائه می‌دهد. همچنین سازوکارهای به‌روزرسانی خودکار و یکپارچگی با ابزارهای متن‌باز (مانند Elasticsearch، Airflow و Kibana) نشان می‌دهند که این رویکرد می‌تواند به‌عنوان یک راهکار عملی و مقیاس‌پذیر برای بایوبانک‌های داده‌های تصویربرداری مغزی مورد استفاده قرار گیرد.

۳-۶ جمع بندی

در این فصل ابتدا نیازمندی های اصلی سیستم و معماری پیشنهادی برای مدیریت و بازیابی داده های تصویربرداری مغزی مبتنی بر استاندارد BIDS معرفی شد. سپس مراحل مختلف پیاده سازی شامل استخراج و پردازش داده ها، ایندکس گذاری و ذخیره سازی در Elasticsearch، پایش و مصورسازی داده ها در Kibana، طراحی رابط کاربری برای افزودن داده های جدید، ماژول به روزرسانی خودکار با Airflow و در نهایت یکپارچه سازی ماژول ها در بستر Docker تشریح گردید.

در ادامه، روش ارزیابی بر مبنای یک مجموعه داده واقعی و سه سناریوی عملی شامل افزودن داده جدید، جستجو و بازیابی اطلاعات، و پایداری به روزرسانی های دوره ای طراحی و اجرا شد. نتایج نشان داد که سامانه ی پیشنهادی از نظر پویایی در مدیریت داده ها، سرعت و کارایی در جستجو و همچنین قابلیت اطمینان در به روزرسانی های خودکار به اهداف تعریف شده دست یافته است.

به طور کلی، یافته های این فصل بیانگر آن است که رویکرد ارائه شده توانسته است نیازمندی های اولیه پروژه را پوشش دهد و زیرساختی یکپارچه، مقیاس پذیر و قابل اتکا برای مدیریت داده های پیچیده تصویربرداری مغزی فراهم سازد.

- [١] Van Essen DC, Smith SM, Barch DM, Behrens TE, Yacoub E, Ugurbil K; WU-Minn HCP Consortium. The WU-Minn Human Connectome Project: an overview. *Neuroimage*. 2013 Oct 15;80:62-79. doi: 10.1016/j.neuroimage.2013.05.041. Epub 2013 May 16. PMID: 23684880; PMCID: PMC3724347.
- [٢] Elam, Jennifer & Glasser, Matthew & Harms, Michael & Sotiropoulos, Stamatis & Andersson, Jesper & Burgess, Gregory & Curtiss, Sandra & Oostenveld, Robert & Larson-Prior, Linda & Schoffelen, Jan-Mathijs & Hodge, Michael & Cler, Eileen & Marcus, Daniel & Barch, Deanna & Yacoub, Essa & Smith, Stephen & Ugurbil, Kamil & Van Essen, David. (2021). The Human Connectome Project: A Retrospective. *NeuroImage*. 244. 118543. 10.1016/j.neuroimage.2021.118543.
- [٣] Littlejohns, T.J., Holliday, J., Gibson, L.M. et al. The UK Biobank imaging enhancement of 100,000 participants: rationale, data collection, management and future directions. *Nat Commun* 11, 2624 (2020). <https://doi.org/10.1038/s41467-020-15948-9>
- [٤] Gorgolewski, K., Auer, T., Calhoun, V. et al. The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Sci Data* 3, 160044 (2016). <https://doi.org/10.1038/sdata.2016.44>
- [٥] <https://openneuro.org/>
- [٦] Tulay EE, Metin B, Tarhan N, Arkan MK. Multimodal Neuroimaging: Basic Concepts and Classification of Neuropsychiatric Diseases. *Clin EEG Neurosci*. 2019 Jan;50(1):20-33. doi: 10.1177/1550059418782093. Epub 2018 Jun 20. PMID: 29925268.
- [٧] B. He and Z. Liu, "Multimodal Functional Neuroimaging: Integrating Functional MRI and EEG/MEG," in *IEEE Reviews in Biomedical Engineering*, vol. 1, pp. 23-40, 2008, doi: 10.1109/RBME.2008.2008233.
- [٨] Clunie, D. A. (2000). *DICOM Structured Reporting*. PixelMed Publishing.
- [٩] Cox, R.W. & Ashburner, John & Breman, Hester & Fissell, Kate & Haselgrove, C. & Holmes, C.J. & Lancaster, J.L. & Rex, D.E. & Smith, S.M. & Woodward, J.B. & Strother, Stephen. (2004). A (sort of) new image data format standard: NiFTI-1. 10th Annual Meeting of the Organization for Human Brain Mapping. 22.

- [١٠] Poldrack, Russell A., Christopher J. Markiewicz, Stefan Appelhoff, Yoni K. Ashar, Tibor Auer, Sylvain Baillet, Shashank Bansal et al. "The past, present, and future of the brain imaging data structure (BIDS)." *Imaging Neuroscience* 2 (2024): 1-19.
- [١١] Yaroslav O. Halchenko, Mathias Goncalves, Satrajit Ghosh, Pablo Velasco, Matteo Visconti di Oleggio Castello, Taylor Salo, John T. Wodder II, Michael Hanke, Patrick Sadil, Krzysztof Jacek Gorgolewski, Horea-Ioan Ioanas, Chris Rorden, Timothy J. Hendrickson, Michael Dayan, Sean Dae Houlihan, James Kent, Ted Strauss, John Lee, Isaac To, ... David N. Kennedy. (2025). HeuDiConv — flexible DICOM conversion into structured directory layouts (v1.3.3). Zenodo. <https://doi.org/10.5281/zenodo.15080551>
- [١٢] Yarkoni T, Markiewicz CJ, de la Vega A, Gorgolewski KJ, Salo T, Halchenko YO, McNamara Q, DeStasio K, Poline JB, Petrov D, Hayot-Sasson V, Nielson DM, Carlin J, Kiar G, Whitaker K, DuPre E, Wagner A, Tirrell LS, Jas M, Hanke M, Poldrack RA, Esteban O, Appelhoff S, Holdgraf C, Staden I, Thirion B, Kleinschmidt DF, Lee JA, Visconti di Oleggio Castello M, Notter MP, Blair R. PyBIDS: Python tools for BIDS datasets. *J Open Source Softw.* 2019;4(40):1294. doi: 10.21105/joss.01294. Epub 2019 Aug 12. PMID: 32775955; PMCID: PMC7409983.
- [١٣] Li X, Morgan PS, Ashburner J, Smith J, Rorden C (2016) The first step for neuroimaging data analysis: DICOM to NIfTI conversion. *J Neurosci Methods.* 264:47-56. doi: 10.1016/j.jneumeth.2016.03.001. PMID: 26945974
- [١٤] <https://bids-standard.github.io/pybids/>
- [١٥] <https://github.com/bids-standard/bids-validator>
- [١٦] Pernet, Cyril & Appelhoff, Stefan & Gorgolewski, Krzysztof & Flandin, Guillaume & Phillips, Christophe & Delorme, Arnaud & Oostenveld, Robert. (2019). EEG-BIDS, an extension to the brain imaging data structure for electroencephalography. *Scientific Data.* 6. 10.1038/s41597-019-0104-8.
- [١٧] Younghwa Cha, Yeji Lee, Eunhee Ji, SoHyun Han, Sunhyun Min, Hyoungkyu Kim, Minseo Cho, Haesung Lee, Youngjai Park, and Joon-Young Moon (2025). Sustained Attention Task (gradCPT) Dataset using simultaneous EEG-fMRI and DTI. *OpenNeuro*. [Dataset] doi: doi:10.18112/openneuro.ds006040.v1.0.0

واژه‌نامه

معادل انگلیسی	واژه فارسی
Subject	آزمودنی
Extract	استخراج
Deploy	استقرار
Validation	اعتبارسنجی
Heuristic	الگوریتم ابتکاری
Checkerboard	الگوی شطرنجی
Indexing	ایندکس گذاری
Retrieve	بازیابی
Biobank	بانک زیستی
Stability	پایداری
Monitoring	پایش
Data Processing	پردازش داده
Query	پرس و جو
Dynamism	پویایی
Imagery	تجسم ذهنی
Neuroimaging	تصویربرداری عصبی
Data Flow	جریان داده
Multimodal	چندوجهی
Dashboard	داشبورد
Scheduling	زمانبندی
Infrastructure	زیرساخت
Brain Imaging Data Structure	ساختار داده‌های تصویربرداری مغز

Metadata	فراداده
Workflow	گردش کار
Dataset	مجموعه داده
Modality	مدالیت
Visualization	مصورسازی
Scalability	مقیاس پذیری
Search Engine	موتور جستجو
Integrated	یکپارچه

پیوست ها و ضمائم

Project repository: <https://github.com/MobinaShahbazi/BIDS-ETL>

An Integrated Approach for Storing and Monitoring Neuroimaging Data within the BIDS Framework

Abstract

Recent advances in neuroimaging technologies have led to an exponential increase in multimodal brain data, creating an urgent demand for scalable infrastructures that enable efficient management, indexing, and visualization. The Brain Imaging Data Structure (BIDS) was introduced to standardize the organization of neuroimaging data, yet in large-scale biobanks, challenges such as fast retrieval, interactive visualization, and dynamic updates still remain.

This project presents a proof-of-concept architecture for integrated management of BIDS datasets. Metadata are extracted using PyBIDS, indexed in Elasticsearch, and visualized through interactive dashboards in Kibana. To ensure system dynamism and adaptability to real-world biobank conditions, indexing and updating workflows were automated using Apache Airflow, while all services were deployed in a unified Docker-based infrastructure.

The system was evaluated under three scenarios: (1) adding new data and validating the propagation of changes, (2) comparing PyBIDS and Elasticsearch in terms of retrieval speed and accuracy, and (3) testing the stability of automated periodic updates. Results demonstrated successful reflection of new entries, a retrieval speed approximately 830 times faster with Elasticsearch compared to PyBIDS, and reliable performance of periodic updates, contingent on stable deployment environments.

Overall, the proposed approach effectively addressed the core requirements of the project and provides a scalable, practical solution for managing complex neuroimaging data in biobanks.



Shahid Beheshti University
Faculty of Computer Science and Engineering

An Integrated Approach for Storing and Monitoring Neuroimaging Data within the BIDS Framework

By:
Mobina Shahbazi

A THESIS SUBMITTED
FOR THE DEGREE OF
BACHELOR OF SCIENCE

Supervisor
Dr. Hasan Haghighi

September 2025