

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 پژوهه کارشناسی

دانشجو:

مبینا شهبازی

استاد راهنما:

دکتر حسن حقیقی

الله اذْعُونَكُمْ

سپاسگزاری

سپاس و قدردانی خود را نخست تقدیم می‌کنم به استاد ارجمند، 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 فرآداده‌ها 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 گردش‌کارها. 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 به عنوان یک راهکار مقیاس‌پذیر و عملی برای مدیریت داده‌های پیچیده تصویربرداری مغزی مطرح گردد.

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

^۱ Elasticsearch

فهرست مطالب

۱۰	فصل اول: کلیات
۱۱	۱-۱ مقدمه
۱۲	۱-۲ بیان مسئله
۱۳	۱-۳ کلیات روش پیشنهادی
۱۴	۱-۴ ساختار گزارش
۱۵	فصل دوم: مفاهیم پایه و کارهای مرتبط
۱۶	۱-۲ مقدمه
۱۷	۲-۱ مفاهیم پایه
۱۸	۲-۲-۱ داده‌های تصویربرداری مغز و انواع آن
۱۹	۲-۲-۲ فرمت‌ها و استانداردهای ذخیره‌سازی داده‌ها
۲۰	۲-۲-۳ ابزارها و کتابخانه‌های مرتبط با BIDS
۲۱	۲-۲-۴ ایندکس‌گذاری و جستجو در داده‌ها
۲۲	۳-۱ تحلیل نقاط قوت و ضعف کارهای مشابه
۲۳	۳-۲ جمع‌بندی
۲۴	۴-۱ مقدمه
۲۵	۴-۲ ساختار روش پیشنهادی
۲۶	۴-۳ نیازمندی‌های سیستم
۲۷	۴-۴ نمای کلی سیستم و معماری پیشنهادی
۲۸	۴-۵ فازبندی اجرای پروژه
۲۹	۴-۶ جریان داده‌ها در سیستم
۳۰	۴-۷ پیاده‌سازی روش پیشنهادی
۳۱	۴-۸ استخراج و پردازش داده و فرآداده
۳۲	۴-۹ ایندکس‌گذاری و ذخیره‌سازی
۳۳	۴-۱۰ پایش و مصورسازی داده‌ها

۴۱	رابط کاربری برای افروzen داده‌های جدید	۴-۳-۳
۴۲	۵-۳-۳ مازول به روزرسانی خودکار با Airflow	
۴۳	۶-۳-۳ یکپارچه‌سازی مازول‌ها در زیرساخت Docker	
۴۴	۴-۳ روش ارزیابی	
۴۴	۱-۴-۳ نوع داده‌ها و مجموعه‌داده 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 م DALIته 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 استخراج شده 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- مجموعه داده 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- مجموعه داده 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 Elasticsearch و PyBIDS 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: کلیات

۱-۱ مقدمه

در دهه‌های اخیر، پیشرفت فناوری‌های تصویربرداری عصبی، امکان تولید حجم عظیمی از داده‌های مغزی را فراهم کرده است. داده‌هایی که از مدل‌ایته‌های مختلف مانند MEG، fMRI، EEG 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 پردازش شده به حدود یک پتabyت برسد [۱]، اما تا به امروز، بیش از ۲۷ پتabyت داده به اشتراک گذاشته شده است [۲] همچنین، بانک اطلاعاتی عظیمی مانند بانک زیستی بریتانیا^۳ در مجموع بیش از ۳۰ پتabyت داده شامل داده‌های تصویربرداری، ژنتیکی و سلامت دارد و این حجم به طور 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 به عنوان استاندارد اصلی ذخیره‌سازی داده‌های تصویربرداری پزشکی استفاده می‌شده؛ 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. به عنوان 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 زمان برخواهد 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 به روزرسانی پویا و نمایش گرافیکی اطلاعات کلیدی بایوبانک را داشته باشد.

² Human Connectome Project (HCP)

³ UK Biobank

⁴ Brain Imaging Data Structure (BIDS)

⁵ Metadata

۱-۲ بیان مسئله

با رشد چشمگیر پروژه‌های بزرگ تصویربرداری مغز، نیاز به مدیریت، ذخیره‌سازی و جستجوی کارآمد داده‌های حجمی چندوجهی^۶ بیش از پیش احساس می‌شود. این داده‌ها شامل مطالیته‌های گوناگونی مانند MRI، fMRI، MEG و EEG هستند که هر کدام فرمتهای حجمی 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 مطالیته‌ها یا حجم داده را به صورت تعاملی نمایش دهند.
۳. دشواری نگهداری سیستم‌ها در شرایط داینامیک، که در آن داده‌های جدید به طور 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 به طور کامل در یک سیستم عملیاتی برای مدیریت داده‌های BIDS به کار گرفته است.

بروژه حاضر با هدف ارائه یک اثبات مفهومی^۷، رویکردی ترکیبی برای استخراج فراداده، ایندکس گذاری 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.

⁶ Multi-modal

⁷ Proof of Concept

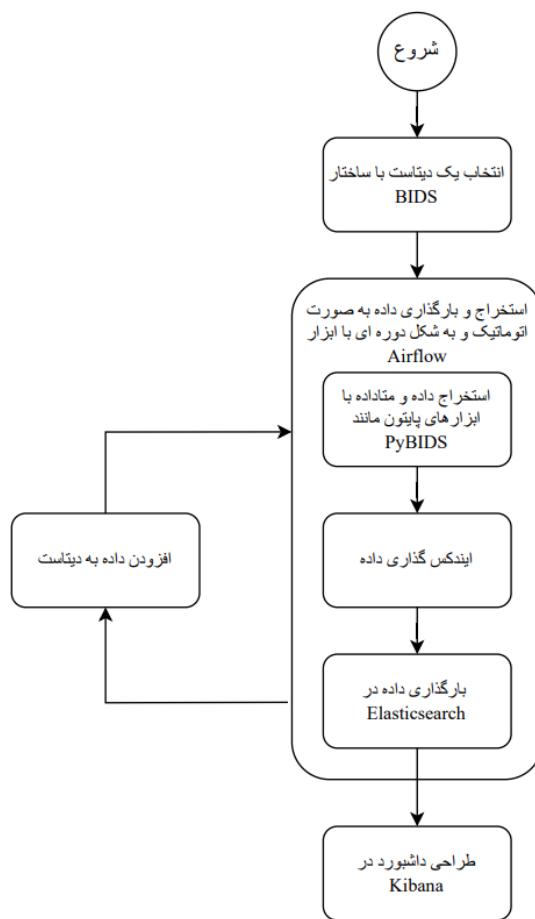
۱-۳ کلیات روش پیشنهادی

در این پژوهه هدف اصلی، طراحی و پیاده‌سازی یک چارچوب یکپارچه برای مدیریت، جستجو و تحلیل فراداده‌های موجود در مجموعه‌داده^۸‌های ساختار BIDS است؛ به گونه‌ای که بتواند در یک محیط پویا مانند بایوبانک داده‌های مغزی، به صورت مستمر به روزرسانی و استفاده شود. در گام نخست، یک مجموعه‌داده 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^۹‌ها، مداریته‌ها، حجم فایل‌ها و محتواهای 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 مداریته‌ها و حجم داده‌ها طراحی می‌گردد. در نهایت، به منظور دینامیک‌سازی سیستم و انطباق با شرایط واقعی بایوبانک، که در آن داده‌های جدید به طور روزانه افزوده می‌شوند، از Apache Airflow برای Docker زمان‌بندی و اجرای خودکار فرآیند استخراج و ایندکس‌گذاری استفاده می‌شود و تمام سرویس‌ها در بستر استقرار می‌یابند. این روش پیشنهادی یک نقشه راه کامل از دریافت داده خام تا ارائه داشبورد تحلیلی و به روزرسانی 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.

⁸ Dataset

⁹ Subject

¹⁰ Index Mapping



شکل ۱- نمای کلی فرایند سیستم

۴-۱ ساختار گزارش

پس از آشنایی با مقدمات موضوع و درک دقیق‌تر 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، فرمتهای و استانداردهای ذخیره‌سازی (DICOM، NIfTI و BIDS)، ابزارها و کتابخانه‌های پشتیبان پردازش و سازماندهی 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، جستجو 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

۱-۲-۱ 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 بعد 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 کامپیوتربازی^{۱۴} نیز اطلاعات ساختاری مفیدی فراهم می‌کنند. از طرف دیگر در حوزه روش‌های عملکردی، تصویربرداری تشخیصی 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 طیف‌سنجدی 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 پایین 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.

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 به تنها‌یی قادر به ارائه تصویر جامع از مغز نیست؛ اما ترکیب آن‌ها نمایی چندبعدی و دقیق‌تر از ساختار و عملکرد مغز فراهم می‌کند. این رویکرد نه تنها 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 روبروست. 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)

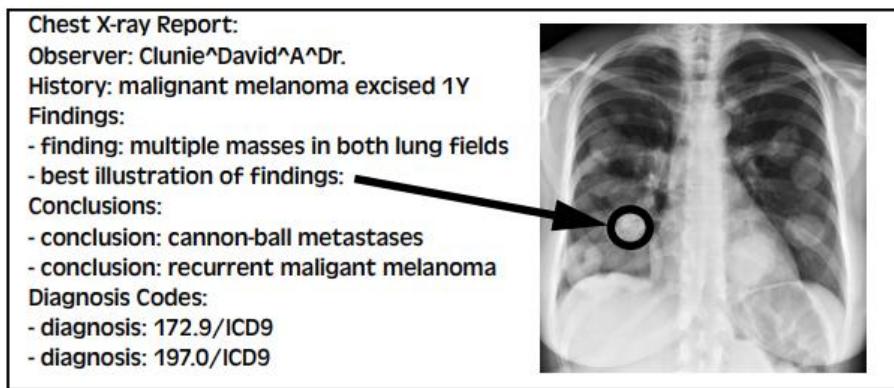
²⁰ Coregistration

روش‌هایی مانند EEG و fMRI ادغام را پیچیده‌تر می‌کند [۶]. نبود استانداردهای یکپارچه برای قالب‌بندی و ثبت داده‌ها نیز مانع 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 DICOM فرمت

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)

NIfTI ۲-۲-۲-۲ فرمت

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 ۱-۱ می‌تواند به صورت یک فایل واحد (.nii) یا دو فایل جدا (.hdr و .img) ذخیره شود [۹].

BIDS ۳-۲-۲-۲ استاندارد

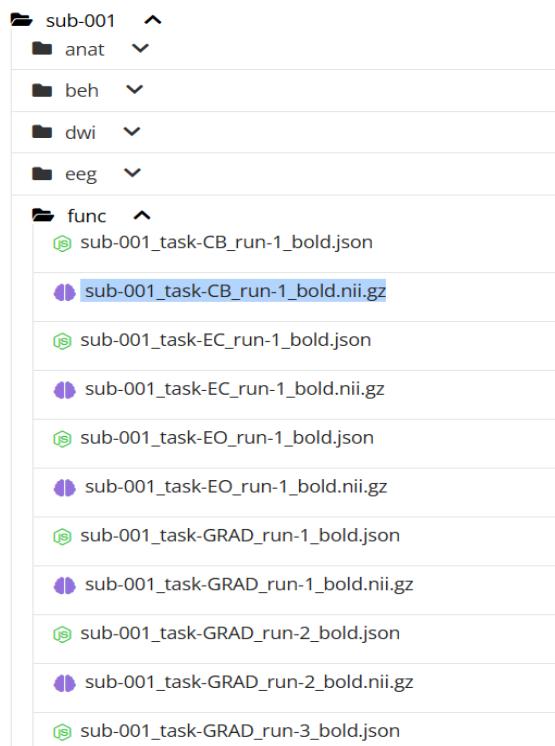
BIDS یک استاندارد برای سازماندهی و توصیف داده‌های تصویربرداری مغزی است که با هدف رفع مشکل پراکنده‌گی و نبود توافق عمومی در ساختاردهی داده‌های MRI و دیگر داده‌های علوم اعصاب ایجاد شده است. پیش از معرفی BIDS، هر آزمایشگاه یا پژوهشگر معمولاً داده‌های خود را به صورت منحصر به فرد و گاهی غیرسازگار با نرم‌افزارها 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 پوشش‌ای منظم و استفاده از فرمتهای فایل رایج و سازگار مانند NIfTI برای داده‌های تصویری، JSON برای فراداده‌ها و TSV برای جداول متونی، رویکردی ساده، شهودی و در عین 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 آسان نر داده‌ها میان پژوهشگران، ادغام مجموعه‌داده‌های 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 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 مدل‌الیته (مانند `anat` برای MRI ساختاری، `func` برای fMRI عملکردی و...) دسته‌بندی می‌شوند. همچنین نام‌گذاری فایل‌ها ترکیبی از برجسب‌های معنادار است، مانند شناسه‌ی فرد، نوع وظیفه^{۲۲}، شماره اجرای تکرار^{۲۳} و نوع داده^{۲۴} [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

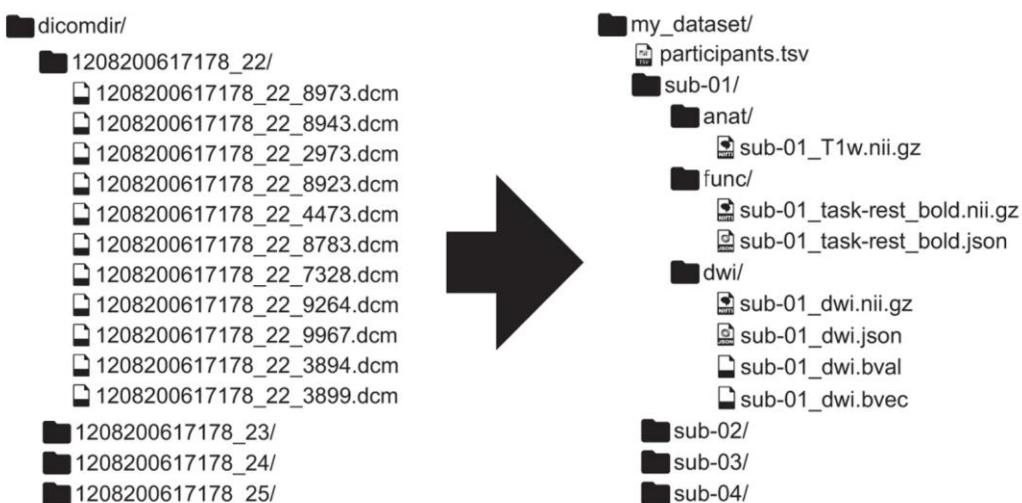
²² Task

²³ Run

²⁴ CheckerBoard

۲-۳-۲-۲ 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 قبل تر 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 dcm2niix 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].

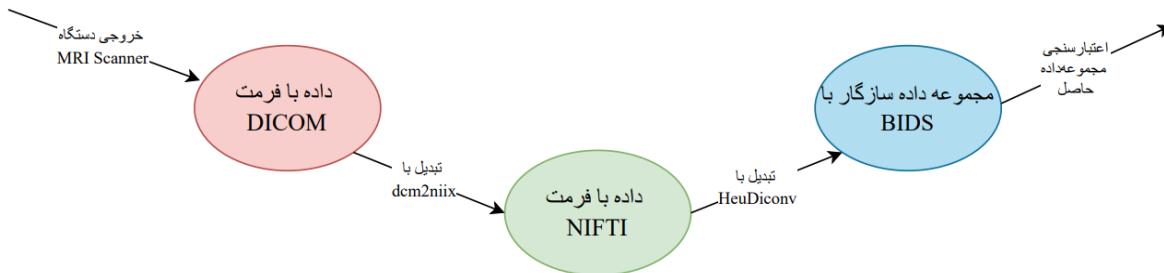


شکل ۴- خروجی‌های آزمایش‌های تصویربرداری عصبی (چپ) و تبدیل آن به مجموعه داده سازگار با BIDS

در گام بعد، فایل‌های NIfTI و فراداده بر اساس ساختار پوشه‌ها و نام‌گذاری استاندارد BIDS سازماندهی می‌شوند. ابزارهایی مانند HeuDiconv²⁵ این کار را با بهره‌گیری از اسکریپت‌های ابتکاری^{۲۶} 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 فرداده اطمینان حاصل شود [۱۱]. سیر تکاملی ذکر شده در شکل ۵ قابل مشاهده است.



شکل ۵- فرایند تبدیل داده خام به قالب BIDS

۲-۳-۲-۲ 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، PyBIDS برای سازماندهی و ساختاردهی، و HeuDiConv برای جستجو، پردازش و تحلیل داده‌های نهایی.

۱-۳-۲-۲ 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 فرداده مهم و استانداردشده است. سادگی و سرعت بالا، پشتیبانی چندسکویی²⁷ و رفع ناسازگاری‌های ناشی از تفاوت پیاده‌سازی DICOM توسط سازندگان مختلف از ویژگی‌های آن است [۱۳].

²⁶ BIDS Validator

²⁷ Cross-platform

HeuDiConv ۲-۳-۲-۲

ابزاری انعطاف‌پذیر برای تبدیل و سازمان‌دهی داده‌های DICOM به ساختار پوشش‌ها 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].

PyBIDS ۳-۳-۲-۲

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 تجمعیع فراداده پردازندۀ 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 به سادگی نصب می‌شود و تعامل با آن عمده‌اً از طریق شیء BIDSLayout انجام می‌گیرد [۱۲][۱۴]. برای مثال کد موجود در شکل ۶ همه فراداده‌های فایل‌های یک مجموعه‌داده 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

۴-۲-۲ ایندکس‌گذاری و جستجو در داده‌ها

۱-۴-۲-۲ مفهوم ایندکس‌گذاری

ایندکس‌گذاری فرایندی است که طی آن داده‌ها به شکلی سازمان‌یافته و بهینه ذخیره می‌شوند تا بتوان آن‌ها را سریع و کارآمد بازیابی کرد. به جای جستجوی مستقیم در کل مجموعه داده که می‌تواند زمان بر و پرهزینه باشد، اطلاعات کلیدی هر رکورد در ساختارهای مانند جداول هش یا درخت‌های جستجو ذخیره می‌شود. این ساختارها که ایندکس نام دارند، همانند فهرست یک کتاب، به سیستم کمک می‌کنند تا محل دقیق داده‌های مورد نظر را با کمترین زمان و منابع بیابد. در حوزه داده‌های علمی و حجیم، ایندکس‌گذاری هم سرعت جستجو را بهبود می‌بخشد، هم امکان انجام پرس‌وجو^{۲۹}‌های پیچیده و فیلترگذاری چندمرحله‌ای را نیز فراهم می‌کند.

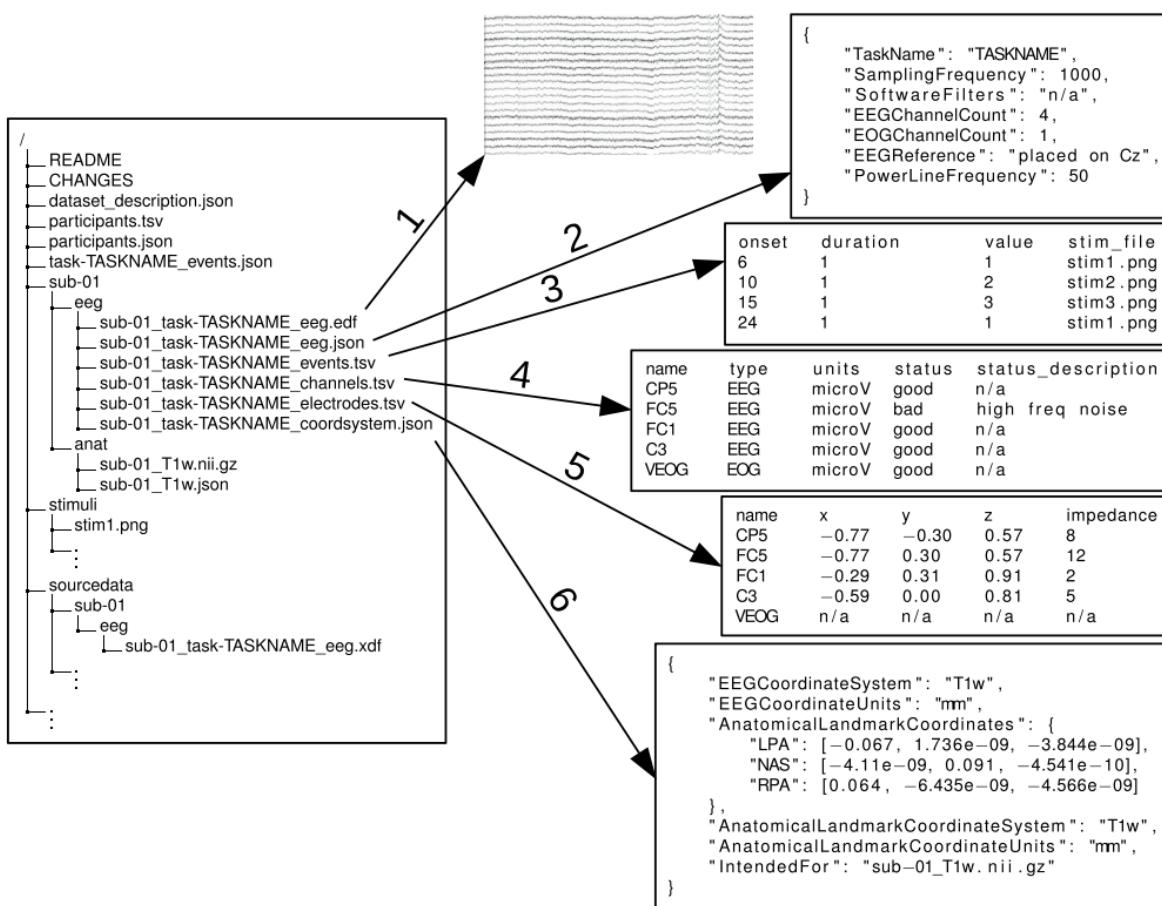
۲-۴-۲-۲ موتور جستجوی Elasticsearch

یک موتور جستجوی متن‌باز است که برای ایندکس‌گذاری و جستجوی سریع داده‌ها بهویژه Elasticsearch داده‌های متنی و ساخت‌یافته در قالب JSON به کار می‌رود. این ابزار قادر است داده‌ها را به صورت اسناد ذخیره و آن‌ها را در شاخص‌های بهینه‌سازی شده سازمان‌دهی کند، به‌گونه‌ای که جستجو حتی در مجموعه‌های داده چند ترابایتی نیز تقریباً به صورت آنی انجام شود. در این پروژه، از Elasticsearch به‌دلیل توانایی بالای آن در مدیریت داده‌های حجیم، پشتیبانی از جستجوهای پیشرفته مانند جستجوی مبتنی بر شباهت یا فیلترگذاری چندلایه و مقیاس‌پذیری بالا استفاده کردہ‌ایم تا بتوان داده‌های تصویربرداری مغزی و فراداده‌های 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 مدلیتهای 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 MEG، EEG و PET ارائه گردید [۱۶].



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

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

و بسیاری از تفاوت‌های ناشی از پیاده‌سازی‌های متفاوت سازندگان اسکنر را نرمال‌سازی می‌نماید [۱۲]. ابزار HeuDiConv نیز به عنوان 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} به عنوان یک وب‌اپلیکیشن، ابزار خط فرمان و کتابخانه جاوا‌اسکریپت/تاپ‌اسکریپت^{۳۱} توسعه یافته تا انطباق داده‌ها با استاندارد BIDS را ارزیابی کند. نسخه وب این ابزار به صورت 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 به عنوان 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 به تنها بی‌فرآیند آماده‌سازی اولیه داده‌ها را انجام نمی‌دهد و بیشتر به عنوان 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 BIDS Validator و dcm2niix، HeuDiConv 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 اضافه شدن مستمر داده‌های جدید و تنوع مدلاتی‌ها 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 روش‌تر 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

۱-۳ 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 به صورت 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 به صورت فازیندی شده توضیح داده خواهد شد و نحوه‌ی جریان داده‌ها در سامانه تبیین می‌گردد. در ادامه، اجزای مختلف پیاده‌سازی شامل مازول‌های پردازش و ایندکس‌گذاری، پایش و مصورسازی، رابط کاربری، به روزرسانی 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 خواهد گرفت.

۲-۳ 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. به منظور خودکارسازی عملیات، فرآیندهای ایندکس‌گذاری و به روزرسانی داده‌ها از طریق گردش کارهای Airflow زمان‌بندی و اجرا می‌شوند.

³³ Multi-modal

۱-۲-۳ 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 مجموعه‌داده
- امکان فیلتر کردن دادگان بر حسب ویژگی‌های مختلف
- امکان افزودن آزمودنی جدید
- مصورسازی دادگان و فراداده‌های مجموعه‌داده و ارائه داشبوردهای تصویری برای پایش و تحلیل اطلاعات کلیدی
- به روز رسانی داشبوردها با تغییر مجموعه‌داده

● نیازمندی‌های غیرعملکردی:

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

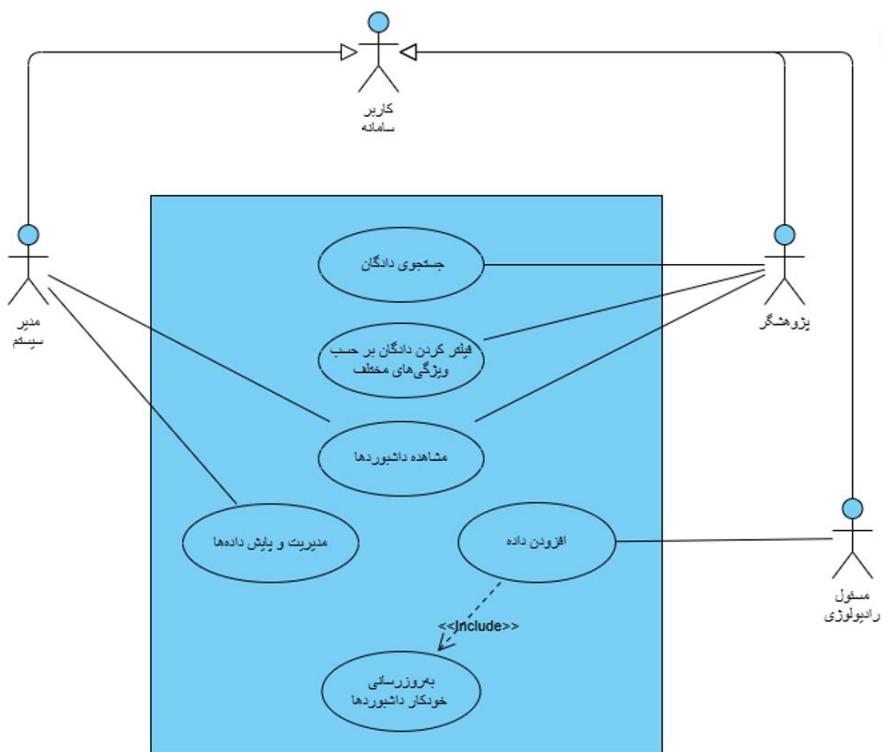
در شکل ۸ نمودار موارد کاربردی^{۳۶} سیستم پیشنهادی نشان داده شده است که تعامل سه نقش اصلی با سامانه را نمایش می‌دهد. پژوهشگر می‌تواند داده‌ها را جستجو و فیلتر کرده و داشبوردهای تحلیلی را مشاهده کند؛ مدیر سیستم وظیفه مدیریت و پایش داده‌ها را بر عهده دارد؛ و مسئول رادیولوژی از طریق رابط کاربری وب امکان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

³⁴ Functional

³⁵ Non-functional

³⁶ Use Case Diagram

فصل سوم: روش پیشنهادی و نتیجه گیری



شکل ۸- نمودار موارد کاربری

۲-۲-۳ نمای کلی سیستم و معماری پیشنهادی

سیستم پیشنهادی در پنج لایه‌ی اصلی طراحی شده است که هر کدام وظیفه‌ی مشخصی در مدیریت، پردازش و تحلیل داده‌ها بر عهده 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 مجموعه داده 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

فصل سوم: روش پیشنهادی و نتیجه گیری

و فراداده مربوط به آن در فایل TSV ثبت می‌گردد تا امکان جستجو و پردازش دقیق‌تر فراهم شود. این لایه به عنوان 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 پردازش شده 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

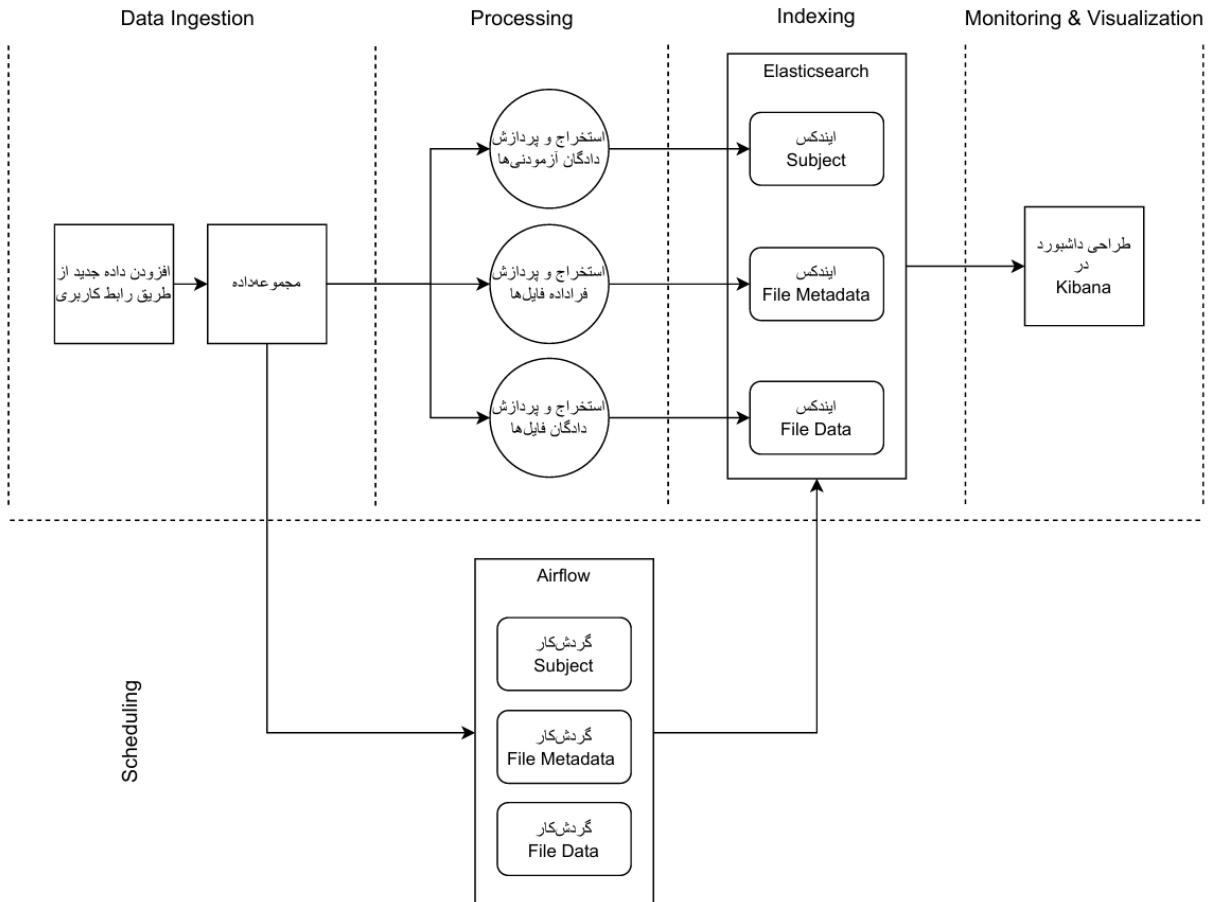
⁴² DAG

⁴³ Monitoring

⁴⁴ Visualization

فصل سوم: روش پیشنهادی و نتیجه گیری

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 پرداختم. 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 استخراج شده 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 گردش‌کار 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.

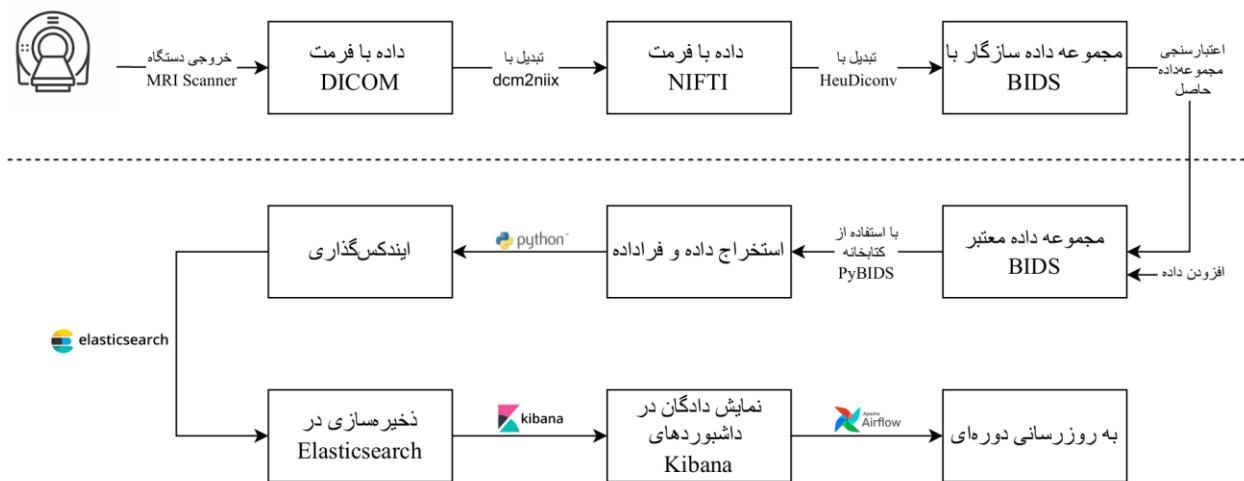
۴-۲-۳ جریان داده‌ها در سیستم

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 گردش‌کار 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.

فصل سوم: روش پیشنهادی و نتیجه گیری



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 به طور دقیق مطالعه شد و کتابخانه PyBIDS به عنوان ابزار اصلی جهت پیمایش و استخراج اطلاعات انتخاب گردید. همچنین به منظور دستیابی به مجموعه داده‌های معتبر، از پایگاه OpenNeuro استفاده شد که دسترسی به مجموعه داده‌های 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 مداریت‌های هر آزمودنی آماده گردید.

فصل سوم: روش پیشنهادی و نتیجه گیری

- در اسکریپت دوم، تمرکز بر سطح فایل‌ها بود و فراداده‌ای همچون مسیر فایل، اندازه‌ی فایل و نوع مدالیته 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؛ به‌گونه‌ای که برای مدالیته‌های *dwi* و *anat*، فراداده‌ای مرتبط از فایل‌های JSON (مانند قدرت میدان مغناطیسی، زمان تکرار، زمان اکو، زاویه چرخش، نوع تصویر و اطلاعات دستگاه MRI) استخراج گردید. برای مدالیته‌ی *beh* نیز اطلاعات رفتاری موجود در فایل‌های TSV مورد پردازش قرار گرفت.

این سه لایه پردازشی مکمل یکدیگر بوده و داده‌های استخراج شده 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

۲-۳-۳ ایندکس‌گذاری و ذخیره‌سازی

در گام بعد، داده‌ها و فراداده‌های استخراج شده در قالب ایندکس‌های مستقل در Elasticsearch ذخیره سازی شدند. برای این منظور ابتدا سرویس Elasticsearch در محیط Docker مستقر گردید که انعطاف‌پذیری و سهولت مدیریت زیرساخت را تضمین می‌کرد. پس از بررسی و تست اتصال از طریق ابزار Postman، عملیات ایندکس‌گذاری از درون اسکریپت‌های پایتونی اجرا شد (شکل ۱۲).

هر یک از اسکریپت‌های سه‌گانه داده‌های آماده شده را به ایندکس منتظر خود منتقل کردند: ایندکس آزمودنی‌ها (subject)، ایندکس فایل‌ها (file)، ایندکس فراداده‌های مربوط به مدلیته‌های تصویری (anat, func, dwi) و ایندکس مربوط به داده‌های رفتاری (beh). بدین ترتیب ساختاری یکپارچه و جستجو محور شکل گرفت که امکان واکشی سریع، انعطاف‌پذیر و هدفمند داده‌ها را در مراحل بعدی تحلیل و مصورسازی فراهم می‌کرد.

```
es = Elasticsearch(  
    hosts=[{  
        'host': 'localhost',  
        'port': 9200,  
        'scheme': 'http'  
    }],  
    basic_auth=('elastic', '12345678')  
)  
  
# 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}.")
```

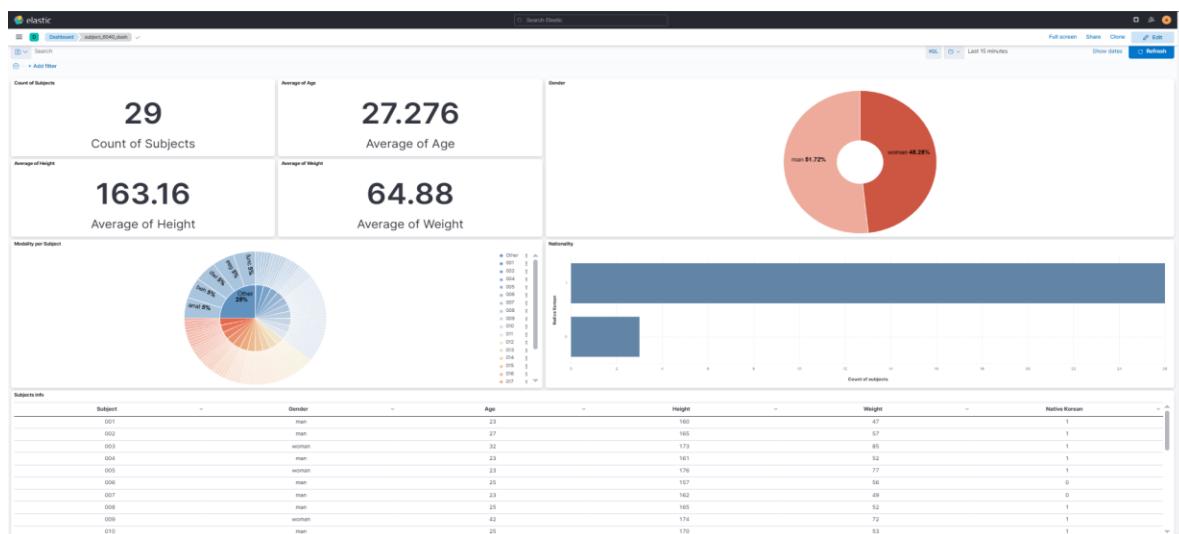
شکل ۱۲- انتقال دادگان استخراج شده به Elasticsearch

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 به صورت یک الگوی شاخص^{۴۵} تعریف شد تا بتواند به عنوان 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 مدالیته‌ها را مصورسازی کرد. داشبورد سوم نیز بر محتوای فایل‌ها متمرکز بوده و شاخص‌های کلیدی استخراج شده از فراداده‌های تصویری (مانند زمان تکرار، زمان اکو، ضخامت برش و نوع تصویر) را به صورت 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 جزئی تر میان آزمودنی‌ها و مدالیته‌ها را مهیا می‌کند. (شکل‌های ۱۳ تا ۱۵)

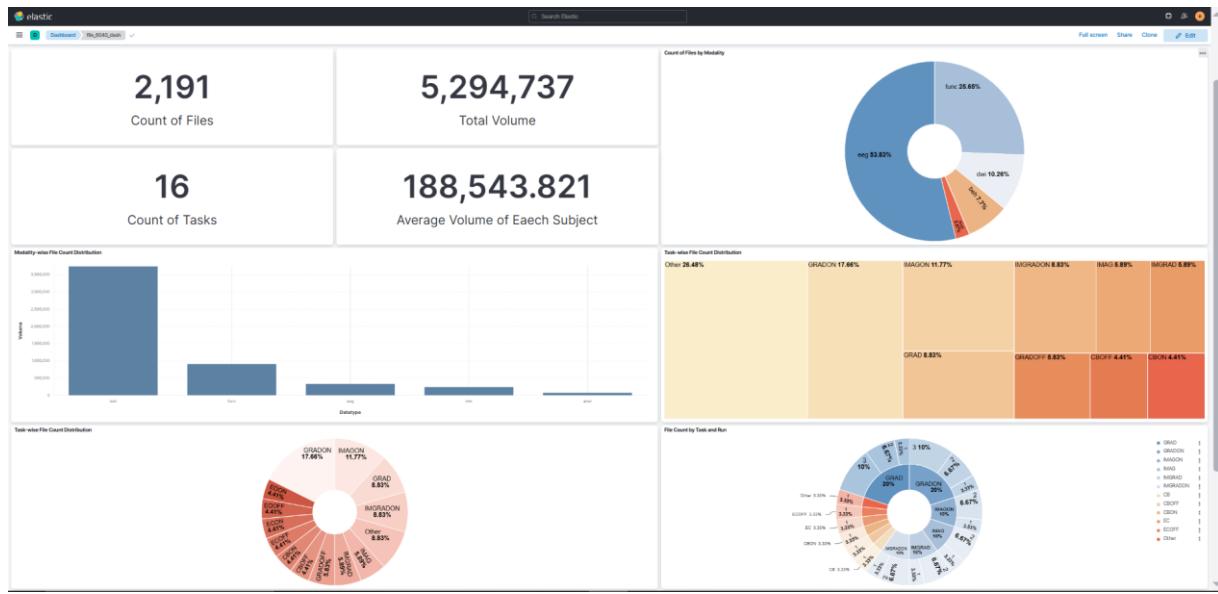


شکل ۱۳ - نمای داشبورد نخست (اطلاعات دموگرافیک آزمودنی‌ها)

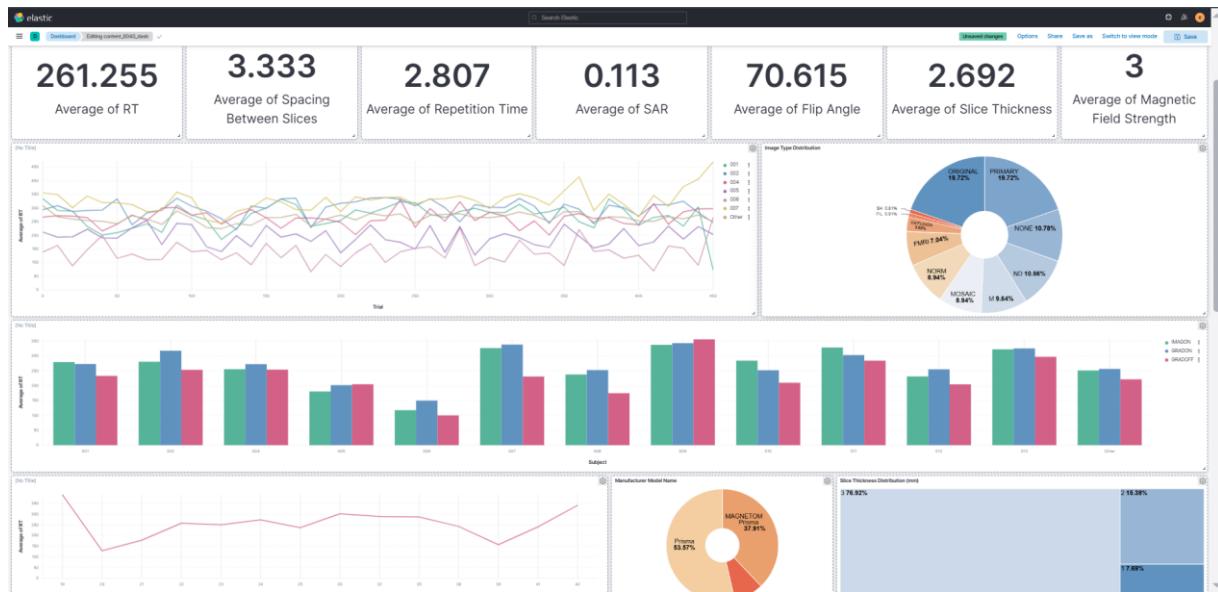
⁴⁵ Index Pattern

⁴⁶ metric

فصل سوم: روش پیشنهادی و نتیجه گیری



شكل ۱۴ - نمای داشبورد دوم (فراداده فایل‌ها)

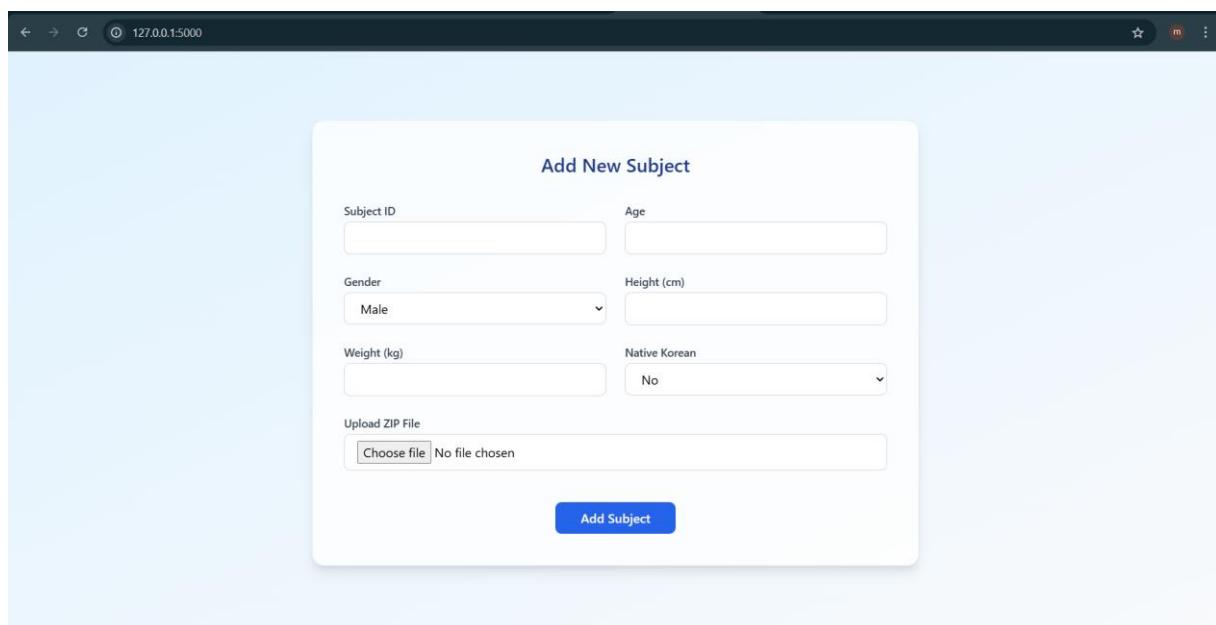


شكل ۱۵ - نمای داشبورد سوم (محتوای فایل‌ها)

۴-۳-۳ 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 تكمیلی) 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 وارد شده به همراه 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 به کاررفته در مراحل پیشین (نظیر PyBIDS و اسکریپت‌های ایندکس‌گذاری) برای پردازش داده‌های جدید نیز استفاده کرد. در شکل ۱۶ نمونه‌ای از رابط کاربری طراحی شده برای ثبت آزمودنی‌های جدید نشان داده شده است.



The screenshot shows a web application interface for adding a new subject. The title of the modal window is "Add New Subject". The form contains the following fields:

- Subject ID: An input field.
- Age: An input field.
- Gender: A dropdown menu set to "Male".
- Height (cm): An input field.
- Weight (kg): An input field.
- Native Korean: A dropdown menu set to "No".
- Upload ZIP File: A section with a "Choose file" button and a message "No file chosen".
- Add Subject: A blue button at the bottom of the form.

شکل ۱۶ - رابط کاربری برای افزودن داده

۵-۳-۳ 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 گردش‌کار مستقل در Airflow پیاده‌سازی شدند. هر گردش‌کار وظیفه‌ی اجرای روزانه‌ی یک بخش از فرآیند ایندکس‌گذاری (آزمودنی‌ها، فایل‌ها و فراداده) را بر عهده دارد. با این مکانیزم، هرگونه تغییر یا داده‌ی افزوده شده 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 گردش‌کارها به همراه سابقه تعداد دفعات اجرای موفق یا ناموفق هر یک نمایش داده می‌شوند. همچنین قابلیت تحریک کردن هر گردش‌کار به شکل دستی در این محیط وجود دارد. برای مثال شکل ۱۷ نشان می‌دهد اولین گردش‌کار 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	3 (2 green, 1 red)	@daily	2025-08-27, 00:00:00	2025-08-28, 00:00:00	5 (all green)	Run	Link
bids_info_6040	airflow	3 (2 green, 1 red)	None	2025-08-12, 16:03:21		5 (all green)	Run	Link
bids_subjects_6040	airflow	3 (2 green, 1 red)	@daily	2025-08-27, 00:00:00	2025-08-28, 00:00:00	5 (all green)	Run	Link

Showing 1-3 of 3 DAGs

Version: v2.8.1
Git Version: release:c0ffa9c5d96625c68ded9562832674ed366b5eb3

شکل ۱۷ - رابط کاربری سرویس Apache Airflow

فصل سوم: روش پیشنهادی و نتیجه گیری

نمونه‌ای از تعریف یک گرددش کار در Airflow برای بارگذاری اطلاعات فایل‌های مجموعه‌داده 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 گرددش کار

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

۶-۳-۶- یکپارچه‌سازی مازول‌ها 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، Worker و Scheduler) 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.

۴-۳ روش ارزیابی

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

برای ارزیابی روش پیشنهادی، از مجموعه‌داده *OpenNeuro – Sustained Attention Task (gradCPT)* 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 Siemens Magnetom ۳ T بازه سیمی می‌باشد. EEG با استفاده از Brain Product BrainCap MR با ۶۴ کانال Prisma جمع‌آوری شده و با استفاده از EEG بازه سیمی می‌باشد. این مجموعه شامل ۴۷ 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 مداریته‌ها، 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 مداریته 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

فصل سوم: روش پیشنهادی و نتیجه گیری

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

۲-۴-۳ سناریوهای ارزیابی عملکرد

به منظور سنجش کارایی زیرساخت طراحی شده، سه سناریوی اصلی تعریف گردید که هرکدام 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 Airflow، Elasticsearch 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 دقیقت در پاسخ به پرس‌وجوها
- پایداری و صحت بهروزرسانی‌های دوره‌ای در فرآیند اجرای خودکار گردش‌کارها و حفظ یکپارچگی داده‌ها

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

۱-۲-۴-۳ سناریو اول، افزودن داده جدید

برای سنجش پویایی سیستم، یک چرخه‌ی کامل از وضعیت اولیه، افزودن داده جدید سپس بهروزرسانی خودکار و نهایتا مشاهده تغییرات در داشبوردها پیاده‌سازی می‌شود. معیار موفقیت این سناریو آن است که افزایش تعداد آزمودنی‌ها در سطح مجموعه‌داده، پس از اجرای گردش‌کار 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:

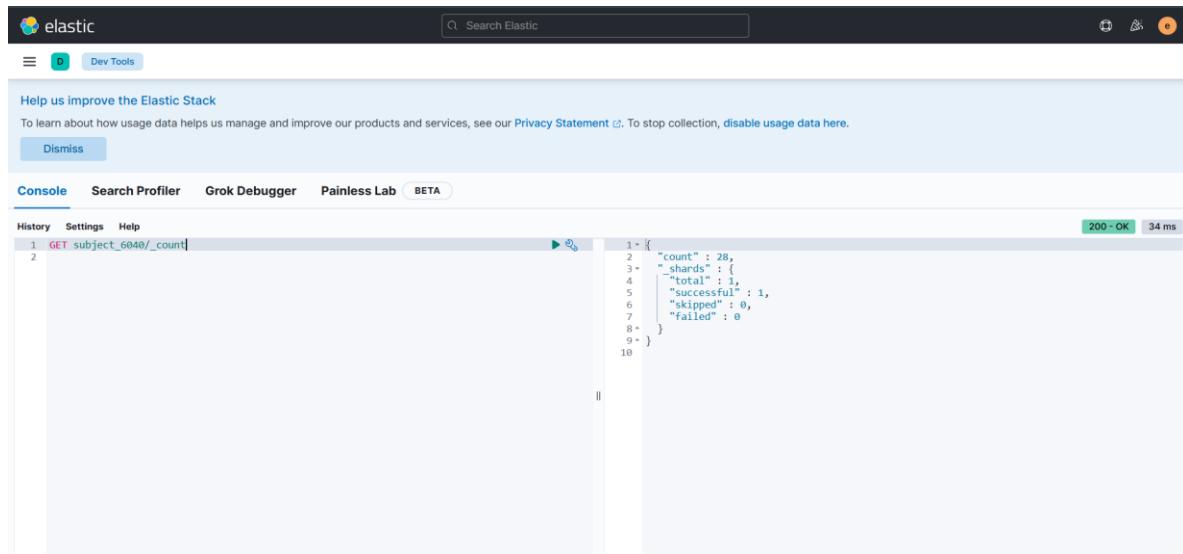
- سطح اول) مجموعه‌داده خام (ساختار پوشش‌ها): همانطور که در شکل ۱۹ مشاهده می‌شود، تعداد آزمودنی‌ها هم در فایل TSV و هم در میان پوشش‌های هر آزمودنی، ۲۸ می‌باشد.

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	

شکل ۱۹ - مجموعه‌داده 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.



The screenshot shows the Elasticsearch Dev Tools interface. In the top navigation bar, 'Console' is selected. Below it, a code editor window displays the following KQL query:

```
1 GET subject_6040/_count
2
```

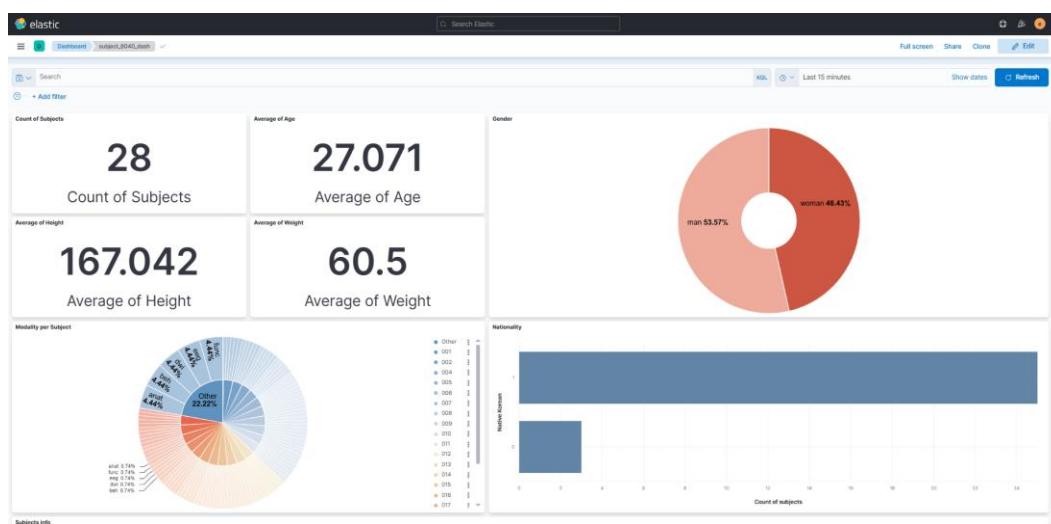
To the right of the code editor, the response is shown in JSON format:

```
1 { "count": 28,
2   "shards": [
3     "total": 1,
4     "successful": 1,
5     "skipped": 0,
6     "failed": 0
7   ]
8 }
9 }
```

At the bottom right of the response area, it says '200 - OK' and '34 ms'.

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 URL 127.0.0.1:5000. Inside, a modal dialog titled "Add New Subject" is displayed. The form contains the following data:

Subject ID	sub-029	Age	33
Gender	Male	Height (cm)	178
Weight (kg)	80	Native Korean	No
Upload ZIP File Choose file: sub-029.zip			

At the bottom of the modal is a blue "Add Subject" button.

شکل ۲۲ - افزودن داده از طریق رابط کاربری

۳. مرحله سوم: بررسی مجدد مجموعه داده

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

فصل سوم: روش پیشنهادی و نتیجه گیری

The screenshot shows a Jupyter Notebook environment. On the left, the 'EXPLORER' sidebar displays a tree structure under 'UNTITLED (WORKSPACE)'. The 'bids-data' folder contains numerous entries labeled 'sub-004' through 'sub-029'. Below this is a file named '.bidsignore'. On the right, a code cell titled 'Demographic_Information.tsv' is open, showing a tab-separated values (TSV) file. The file has 32 rows and 7 columns, with headers: Anonymized ID, Age, Gender, Height, Weight, Native, and Korean. The data includes various demographic information for subjects like sub-001 to sub-029.

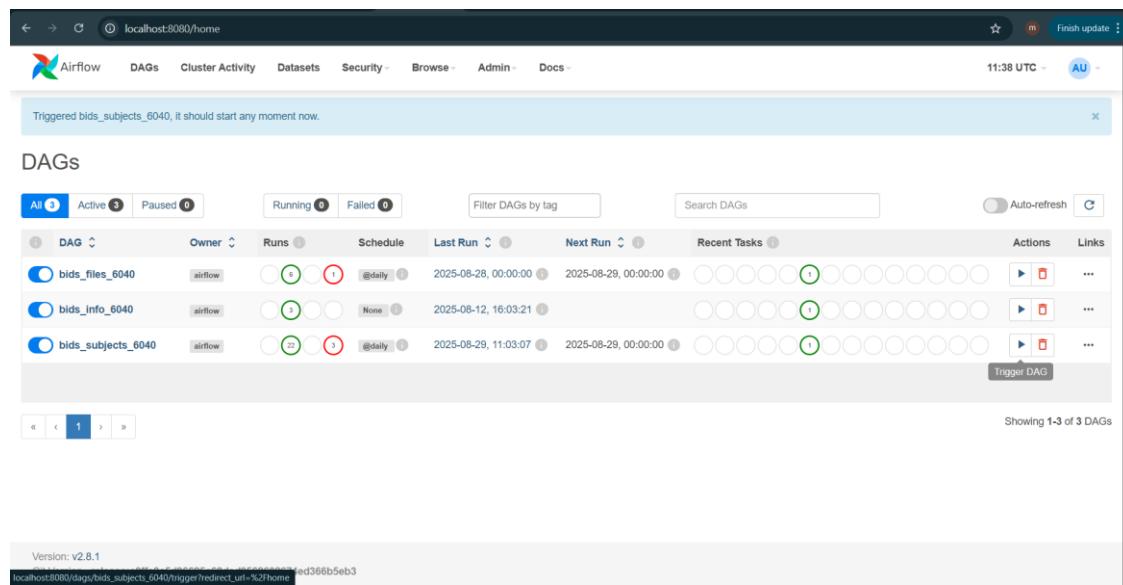
	Anonymized ID	Age	Gender	Height	Weight	Native	Korean
1	ID	Age	Gender	Height	Weight	Native	Korean
2	sub-001	23	1	160	47	1	
3	sub-002	27	1	165	57	1	
4	sub-003	32	0	173	85	1	
5	sub-004	23	1	161	52	1	
6	sub-005	23	0	176	77	1	
7	sub-006	25	1	157	56	0	
8	sub-007	23	1	162	49	0	
9	sub-008	25	1	165	52	1	
10	sub-009	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	sub-029	33	1	178	80	0	
31							
32							

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

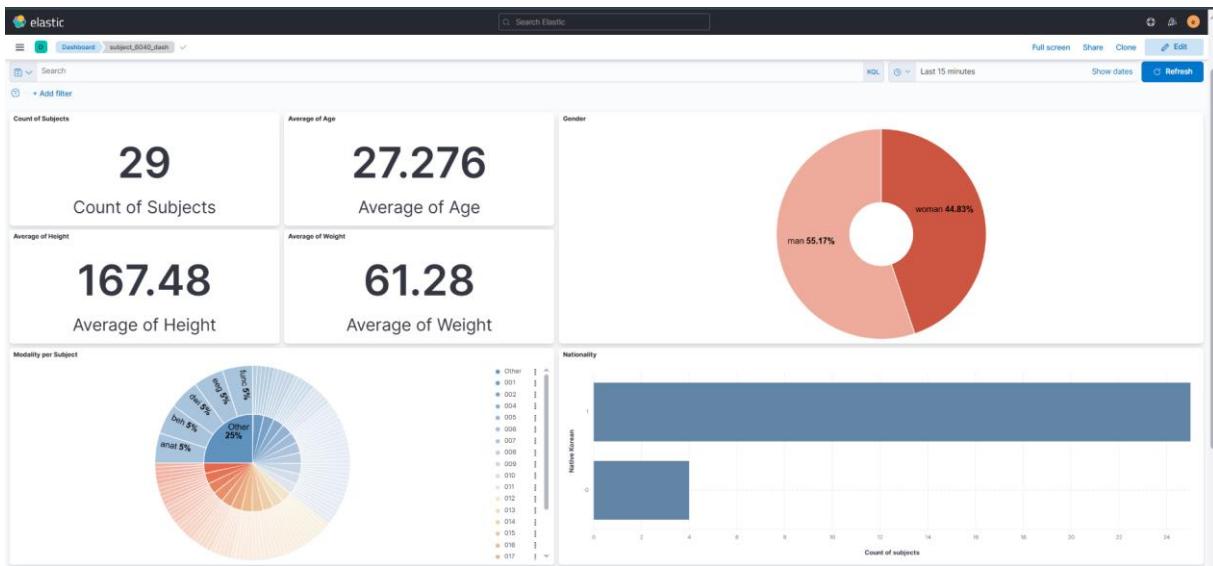
۵. مرحله پنجم: نمایش وضعیت به روز شده در Kibana و Elasticsearch

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 روز شده نشان داده می‌شود (شکل ۲۶).

```
1 GET subject_6040/_count
2
3 {
4     "count": 29
5 }
6
7 }
```

شکل ۲۵ - پرس‌وجو برای شمارش آزمودنی‌ها در ایندکس subject_6040

فصل سوم: روش پیشنهادی و نتیجه گیری



شکل ۲۶ - نمای داشبورد آزمودنی‌ها

۲-۲-۴-۳ 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 آن‌جا 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 پرس‌وحو مشخص همزمان با استفاده از Elasticsearch و PyBIDS 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 پرس‌وحوها 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 استخراج شده نتایج را برمی‌گرداند.

```
import time
from bids import BIDSLayout

start = time.time()

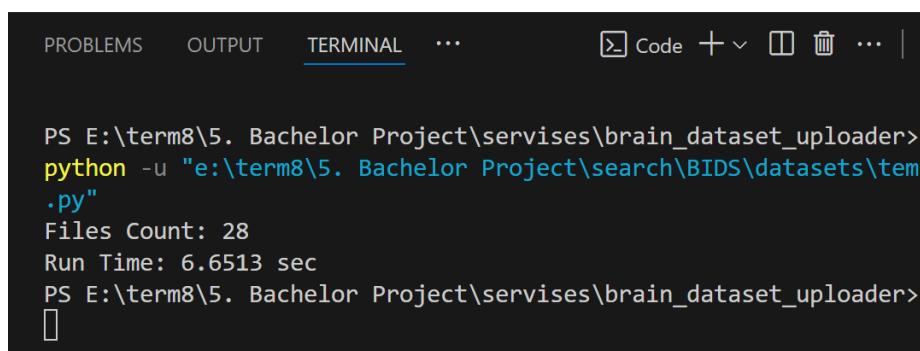
layout = BIDSLayout(r"services\bids-data")
anat_files = layout.get(datatype="anat", extension="nii.gz")

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.



The screenshot shows a terminal window with the following interface elements at the top:

- PROBLEMS
- OUTPUT
- TERMINAL
- ...
- Code
- + ▾
- ✖
- ...
- |

The terminal output is as follows:

```
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

فصل سوم: روش پیشنهادی و نتیجه گیری

۲. جستجو با موتور جستجوی Elasticsearch

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

The screenshot shows the Elasticsearch Dev Tools interface. In the left panel, a code editor displays a GET request to the '_search' endpoint of the 'file_6040' index. The query uses a bool query with a must clause containing two match queries: one for 'datatype: anat' and another for 'extension: nii.gz'. The right panel shows the JSON response. The 'took' field is 8, indicating the search took 8 milliseconds. The 'hits' field shows a total of 28 matches, each with a score of 5.4478025 and a source object containing file details like path, datatype, extension, suffix, and filesize. The status bar at the bottom indicates a 200 OK response with 74 ms latency.

```

1 - GET file_6040/_search
2 + {
3 +   "query": {
4 +     "bool": {
5 +       "must": [
6 +         { "match": { "datatype": "anat" } },
7 +         { "match": { "extension": "nii.gz" } }
8 +       ]
9 +     }
10 +   }
11 + }
12 + }
13 + }
14 + }
15 + }
16 + }
17 + }
18 + }
19 + }
20 + }
21 + }
22 + }
23 + }
24 + }
25 + }
26 + }
27 + }
28 + }
29 + }
30 + }
31 + }

1 - {
2 -   "took" : 8,
3 -   "timed_out" : false,
4 -   "_shards" : {
5 -     "total" : 1,
6 -     "successful" : 1,
7 -     "skipped" : 0,
8 -     "failed" : 0
9 -   }
10 -   "hits" : {
11 -     "total" : {
12 -       "value" : 28,
13 -       "relation" : "eq"
14 -     },
15 -     "max_score" : 5.4478025,
16 -     "hits" : [
17 -       {
18 -         "_index" : "file_6040",
19 -         "_type" : "_doc",
20 -         "_id" : "qnJ992gBvYl_G_HuaPla",
21 -         "_score" : 5.4478025,
22 -         "_source" : {
23 -           "path" : "/opt/airflow/bids-data/sub-001/anat/sub-001_T1w.nii.gz",
24 -           "datatype" : "anat",
25 -           "extension" : ".nii.gz",
26 -           "suffix" : "T1w",
27 -           "filesize" : 210
28 -         }
29 -       }
30 -     ]
31 -   }
32 - }
33 - }
34 - }
35 - }
36 - }
37 - }
38 - }
39 - }
40 - }
41 - }
42 - }
43 - }
44 - }
45 - }
46 - }
47 - }
48 - }
49 - }
50 - }
51 - }
52 - }
53 - }
54 - }
55 - }
56 - }
57 - }
58 - }
59 - }
60 - }
61 - }
62 - }
63 - }
64 - }
65 - }
66 - }
67 - }
68 - }
69 - }
70 - }
71 - }
72 - }
73 - }
74 - }
75 - }
76 - }
77 - }
78 - }
79 - }
80 - }
81 - }
82 - }
83 - }
84 - }
85 - }
86 - }
87 - }
88 - }
89 - }
90 - }
91 - }
92 - }
93 - }
94 - }
95 - }
96 - }
97 - }
98 - }
99 - }
100 - }
101 - }
102 - }
103 - }
104 - }
105 - }
106 - }
107 - }
108 - }
109 - }
110 - }
111 - }
112 - }
113 - }
114 - }
115 - }
116 - }
117 - }
118 - }
119 - }
120 - }
121 - }
122 - }
123 - }
124 - }
125 - }
126 - }
127 - }
128 - }
129 - }
130 - }
131 - }
132 - }
133 - }
134 - }
135 - }
136 - }
137 - }
138 - }
139 - }
140 - }
141 - }
142 - }
143 - }
144 - }
145 - }
146 - }
147 - }
148 - }
149 - }
150 - }
151 - }
152 - }
153 - }
154 - }
155 - }
156 - }
157 - }
158 - }
159 - }
160 - }
161 - }
162 - }
163 - }
164 - }
165 - }
166 - }
167 - }
168 - }
169 - }
170 - }
171 - }
172 - }
173 - }
174 - }
175 - }
176 - }
177 - }
178 - }
179 - }
180 - }
181 - }
182 - }
183 - }
184 - }
185 - }
186 - }
187 - }
188 - }
189 - }
190 - }
191 - }
192 - }
193 - }
194 - }
195 - }
196 - }
197 - }
198 - }
199 - }
200 - }
201 - }
202 - }
203 - }
204 - }
205 - }
206 - }
207 - }
208 - }
209 - }
210 - }
211 - }
212 - }
213 - }
214 - }
215 - }
216 - }
217 - }
218 - }
219 - }
220 - }
221 - }
222 - }
223 - }
224 - }
225 - }
226 - }
227 - }
228 - }
229 - }
230 - }
231 - }
232 - }
233 - }
234 - }
235 - }
236 - }
237 - }
238 - }
239 - }
240 - }
241 - }
242 - }
243 - }
244 - }
245 - }
246 - }
247 - }
248 - }
249 - }
250 - }
251 - }
252 - }
253 - }
254 - }
255 - }
256 - }
257 - }
258 - }
259 - }
260 - }
261 - }
262 - }
263 - }
264 - }
265 - }
266 - }
267 - }
268 - }
269 - }
270 - }
271 - }
272 - }
273 - }
274 - }
275 - }
276 - }
277 - }
278 - }
279 - }
280 - }
281 - }
282 - }
283 - }
284 - }
285 - }
286 - }
287 - }
288 - }
289 - }
290 - }
291 - }
292 - }
293 - }
294 - }
295 - }
296 - }
297 - }
298 - }
299 - }
300 - }
301 - }
302 - }
303 - }
304 - }
305 - }
306 - }
307 - }
308 - }
309 - }
310 - }
311 - }
312 - }
313 - }
314 - }
315 - }
316 - }
317 - }
318 - }
319 - }
320 - }
321 - }
322 - }
323 - }
324 - }
325 - }
326 - }
327 - }
328 - }
329 - }
330 - }
331 - }
332 - }
333 - }
334 - }
335 - }
336 - }
337 - }
338 - }
339 - }
340 - }
341 - }
342 - }
343 - }
344 - }
345 - }
346 - }
347 - }
348 - }
349 - }
350 - }
351 - }
352 - }
353 - }
354 - }
355 - }
356 - }
357 - }
358 - }
359 - }
360 - }
361 - }
362 - }
363 - }
364 - }
365 - }
366 - }
367 - }
368 - }
369 - }
370 - }
371 - }
372 - }
373 - }
374 - }
375 - }
376 - }
377 - }
378 - }
379 - }
380 - }
381 - }
382 - }
383 - }
384 - }
385 - }
386 - }
387 - }
388 - }
389 - }
390 - }
391 - }
392 - }
393 - }
394 - }
395 - }
396 - }
397 - }
398 - }
399 - }
400 - }
401 - }
402 - }
403 - }
404 - }
405 - }
406 - }
407 - }
408 - }
409 - }
410 - }
411 - }
412 - }
413 - }
414 - }
415 - }
416 - }
417 - }
418 - }
419 - }
420 - }
421 - }
422 - }
423 - }
424 - }
425 - }
426 - }
427 - }
428 - }
429 - }
430 - }
431 - }
432 - }
433 - }
434 - }
435 - }
436 - }
437 - }
438 - }
439 - }
440 - }
441 - }
442 - }
443 - }
444 - }
445 - }
446 - }
447 - }
448 - }
449 - }
450 - }
451 - }
452 - }
453 - }
454 - }
455 - }
456 - }
457 - }
458 - }
459 - }
460 - }
461 - }
462 - }
463 - }
464 - }
465 - }
466 - }
467 - }
468 - }
469 - }
470 - }
471 - }
472 - }
473 - }
474 - }
475 - }
476 - }
477 - }
478 - }
479 - }
480 - }
481 - }
482 - }
483 - }
484 - }
485 - }
486 - }
487 - }
488 - }
489 - }
490 - }
491 - }
492 - }
493 - }
494 - }
495 - }
496 - }
497 - }
498 - }
499 - }
500 - }
501 - }
502 - }
503 - }
504 - }
505 - }
506 - }
507 - }
508 - }
509 - }
510 - }
511 - }
512 - }
513 - }
514 - }
515 - }
516 - }
517 - }
518 - }
519 - }
520 - }
521 - }
522 - }
523 - }
524 - }
525 - }
526 - }
527 - }
528 - }
529 - }
530 - }
531 - }
532 - }
533 - }
534 - }
535 - }
536 - }
537 - }
538 - }
539 - }
540 - }
541 - }
542 - }
543 - }
544 - }
545 - }
546 - }
547 - }
548 - }
549 - }
550 - }
551 - }
552 - }
553 - }
554 - }
555 - }
556 - }
557 - }
558 - }
559 - }
560 - }
561 - }
562 - }
563 - }
564 - }
565 - }
566 - }
567 - }
568 - }
569 - }
570 - }
571 - }
572 - }
573 - }
574 - }
575 - }
576 - }
577 - }
578 - }
579 - }
580 - }
581 - }
582 - }
583 - }
584 - }
585 - }
586 - }
587 - }
588 - }
589 - }
590 - }
591 - }
592 - }
593 - }
594 - }
595 - }
596 - }
597 - }
598 - }
599 - }
600 - }
601 - }
602 - }
603 - }
604 - }
605 - }
606 - }
607 - }
608 - }
609 - }
610 - }
611 - }
612 - }
613 - }
614 - }
615 - }
616 - }
617 - }
618 - }
619 - }
620 - }
621 - }
622 - }
623 - }
624 - }
625 - }
626 - }
627 - }
628 - }
629 - }
630 - }
631 - }
632 - }
633 - }
634 - }
635 - }
636 - }
637 - }
638 - }
639 - }
640 - }
641 - }
642 - }
643 - }
644 - }
645 - }
646 - }
647 - }
648 - }
649 - }
650 - }
651 - }
652 - }
653 - }
654 - }
655 - }
656 - }
657 - }
658 - }
659 - }
660 - }
661 - }
662 - }
663 - }
664 - }
665 - }
666 - }
667 - }
668 - }
669 - }
670 - }
671 - }
672 - }
673 - }
674 - }
675 - }
676 - }
677 - }
678 - }
679 - }
680 - }
681 - }
682 - }
683 - }
684 - }
685 - }
686 - }
687 - }
688 - }
689 - }
690 - }
691 - }
692 - }
693 - }
694 - }
695 - }
696 - }
697 - }
698 - }
699 - }
700 - }
701 - }
702 - }
703 - }
704 - }
705 - }
706 - }
707 - }
708 - }
709 - }
710 - }
711 - }
712 - }
713 - }
714 - }
715 - }
716 - }
717 - }
718 - }
719 - }
720 - }
721 - }
722 - }
723 - }
724 - }
725 - }
726 - }
727 - }
728 - }
729 - }
730 - }
731 - }
732 - }
733 - }
734 - }
735 - }
736 - }
737 - }
738 - }
739 - }
740 - }
741 - }
742 - }
743 - }
744 - }
745 - }
746 - }
747 - }
748 - }
749 - }
750 - }
751 - }
752 - }
753 - }
754 - }
755 - }
756 - }
757 - }
758 - }
759 - }
760 - }
761 - }
762 - }
763 - }
764 - }
765 - }
766 - }
767 - }
768 - }
769 - }
770 - }
771 - }
772 - }
773 - }
774 - }
775 - }
776 - }
777 - }
778 - }
779 - }
780 - }
781 - }
782 - }
783 - }
784 - }
785 - }
786 - }
787 - }
788 - }
789 - }
790 - }
791 - }
792 - }
793 - }
794 - }
795 - }
796 - }
797 - }
798 - }
799 - }
800 - }
801 - }
802 - }
803 - }
804 - }
805 - }
806 - }
807 - }
808 - }
809 - }
810 - }
811 - }
812 - }
813 - }
814 - }
815 - }
816 - }
817 - }
818 - }
819 - }
820 - }
821 - }
822 - }
823 - }
824 - }
825 - }
826 - }
827 - }
828 - }
829 - }
830 - }
831 - }
832 - }
833 - }
834 - }
835 - }
836 - }
837 - }
838 - }
839 - }
840 - }
841 - }
842 - }
843 - }
844 - }
845 - }
846 - }
847 - }
848 - }
849 - }
850 - }
851 - }
852 - }
853 - }
854 - }
855 - }
856 - }
857 - }
858 - }
859 - }
860 - }
861 - }
862 - }
863 - }
864 - }
865 - }
866 - }
867 - }
868 - }
869 - }
870 - }
871 - }
872 - }
873 - }
874 - }
875 - }
876 - }
877 - }
878 - }
879 - }
880 - }
881 - }
882 - }
883 - }
884 - }
885 - }
886 - }
887 - }
888 - }
889 - }
890 - }
891 - }
892 - }
893 - }
894 - }
895 - }
896 - }
897 - }
898 - }
899 - }
900 - }
901 - }
902 - }
903 - }
904 - }
905 - }
906 - }
907 - }
908 - }
909 - }
910 - }
911 - }
912 - }
913 - }
914 - }
915 - }
916 - }
917 - }
918 - }
919 - }
920 - }
921 - }
922 - }
923 - }
924 - }
925 - }
926 - }
927 - }
928 - }
929 - }
930 - }
931 - }
932 - }
933 - }
934 - }
935 - }
936 - }
937 - }
938 - }
939 - }
940 - }
941 - }
942 - }
943 - }
944 - }
945 - }
946 - }
947 - }
948 - }
949 - }
950 - }
951 - }
952 - }
953 - }
954 - }
955 - }
956 - }
957 - }
958 - }
959 - }
960 - }
961 - }
962 - }
963 - }
964 - }
965 - }
966 - }
967 - }
968 - }
969 - }
970 - }
971 - }
972 - }
973 - }
974 - }
975 - }
976 - }
977 - }
978 - }
979 - }
980 - }
981 - }
982 - }
983 - }
984 - }
985 - }
986 - }
987 - }
988 - }
989 - }
990 - }
991 - }
992 - }
993 - }
994 - }
995 - }
996 - }
997 - }
998 - }
999 - }
1000 - }
1001 - }
1002 - }
1003 - }
1004 - }
1005 - }
1006 - }
1007 - }
1008 - }
1009 - }
1010 - }
1011 - }
1012 - }
1013 - }
1014 - }
1015 - }
1016 - }
1017 - }
1018 - }
1019 - }
1020 - }
1021 - }
1022 - }
1023 - }
1024 - }
1025 - }
1026 - }
1027 - }
1028 - }
1029 - }
1030 - }
1031 - }
1032 - }
1033 - }
1034 - }
1035 - }
1036 - }
1037 - }
1038 - }
1039 - }
1040 - }
1041 - }
1042 - }
1043 - }
1044 - }
1045 - }
1046 - }
1047 - }
1048 - }
1049 - }
1050 - }
1051 - }
1052 - }
1053 - }
1054 - }
1055 - }
1056 - }
1057 - }
1058 - }
1059 - }
1060 - }
1061 - }
1062 - }
1063 - }
1064 - }
1065 - }
1066 - }
1067 - }
1068 - }
1069 - }
1070 - }
1071 - }
1072 - }
1073 - }
1074 - }
1075 - }
1076 - }
1077 - }
1078 - }
1079 - }
1080 - }
1081 - }
1082 - }
1083 - }
1084 - }
1085 - }
1086 - }
1087 - }
1088 - }
1089 - }
1090 - }
1091 - }
1092 - }
1093 - }
1094 - }
1095 - }
1096 - }
1097 - }
1098 - }
1099 - }
1100 - }
1101 - }
1102 - }
1103 - }
1104 - }
1105 - }
1106 - }
1107 - }
1108 - }
1109 - }
1110 - }
1111 - }
1112 - }
1113 - }
1114 - }
1115 - }
1116 - }
1117 - }
1118 - }
1119 - }
1120 - }
1121 - }
1122 - }
1123 - }
1124 - }
1125 - }
1126 - }
1127 - }
1128 - }
1129 - }
1130 - }
1131 - }
1132 - }
1133 - }
1134 - }
1135 - }
1136 - }
1137 - }
1138 - }
1139 - }
1140 - }
1141 - }
1142 - }
1143 - }
1144 - }
1145 - }
1146 - }
1147 - }
1148 - }
1149 - }
1150 - }
1151 - }
1152 - }
1153 - }
1154 - }
1155 - }
1156 - }
1157 - }
1158 - }
1159 - }
1160 - }
1161 - }
1162 - }
1163 - }
1164 - }
1165 - }
1166 - }
1167 - }
1168 - }
1169 - }
1170 - }
1171 - }
1172 - }
1173 - }
1174 - }
1175 - }
1176 - }
1177 - }
1178 - }
1179 - }
1180 - }
1181 - }
1182 - }
1183 - }
1184 - }
1185 - }
1186 - }
1187 - }
1188 - }
1189 - }
1190 - }
1191 - }
1192 - }
1193 - }
1194 - }
1195 - }
1196 - }
1197 - }
1198 - }
1199 - }
1200 - }
1201 - }
1202 - }
1203 - }
1204 - }
1205 - }
1206 - }
1207 - }
1208 - }
1209 - }
1210 - }
1211 - }
1212 - }
1213 - }
1214 - }
1215 - }
1216 - }
1217 - }
1218 - }
1219 - }
1220 - }
1221 - }
1222 - }
1223 - }
1224 - }
1225 - }
1226 - }
1227 - }
1228 - }
1229 - }
1230 - }
1231 - }
1232 - }
1233 - }
1234 - }
1235 - }
1236 - }
1237 - }
1238 - }
1239 - }
1240 - }
1241 - }
1242 - }
1243 - }
1244 - }
1245 - }
1246 - }
1247 - }
1248 - }
1249 - }
1250 - }
1251 - }
1252 - }
1253 - }
1254 - }
1255 - }
1256 - }
1257 - }
1258 - }
1259 - }
1260 - }
1261 - }
1262 - }
1263 - }
1264 - }
1265 - }
1266 - }
1267 - }
1268 - }
1269 - }
1270 - }
1271 - }
1272 - }
1273 - }
1274 - }
1275 - }
1276 - }
1277 - }
1278 - }
1279 - }
1280 - }
1281 - }
1282 - }
1283 - }
1284 - }
1285 - }
1286 - }
1287 - }
1288 - }
1289 - }
1290 - }
1291 - }
1292 - }
1293 - }
1294 - }
1295 - }
1296 - }
1297 - }
1298 - }
1299 - }
1300 - }
1301 - }
1302 - }
1303 - }
1304 - }
1305 - }
1306 - }
1307 - }
1308 - }
1309 - }
1310 - }
1311 - }
1312 - }
1313 - }
1314 - }
1315 - }
1316 - }
1317 - }
1318 - }
1319 - }
1320 - }
1321 - }
1322 - }
1323 - }
1324 - }
1325 - }
1326 - }
1327 - }
1328 - }
1329 - }
1330 - }
1331 - }
1332 - }
1333 - }
1334 - }
1335 - }
1336 - }
1337 - }
1338 - }
1339 - }
1340 - }
1341 - }
1342 - }
1343 - }
1344 - }
1345 - }
1346 - }
1347 - }
1348 - }
1349 - }
1350 - }
1351 - }
1352 - }
1353 - }
1354 - }
1355 - }
1356 - }
1357 - }
1358 - }
1359 - }
1360 - }
1361 - }
1362 - }
1363 - }
1364 - }
1365 - }
1366 - }
1367 - }
1368 - }
1369 - }
1370 - }
1371 - }
1372 - }
1373 - }
1374 - }
1375 - }
1376 - }
1377 - }
1378 - }
1379 - }
1380 - }
1381 - }
1382 - }
1383 - }
1384 - }
1385 - }
1386 - }
1387 - }
1388 - }
1389 - }
1390 - }
1391 - }
1392 - }
1393 - }
1394 - }
1395 - }
1396 - }
1397 - }
1398 - }
1399 - }
1400 - }
1401 - }
1402 - }
1403 - }
1404 - }
1405 - }
1406 - }
1407 - }
1408 - }
1409 - }
1410 - }
1411 - }
1412 - }
1413 - }
1414 - }
1415 - }
1416 - }
1417 - }
1418 - }
1419 - }
1420 - }
1421 - }
1422 - }
1423 - }
1424 - }
1425 - }
1426 - }
1427 - }
1428 - }
1429 - }
1430 - }
1431 - }
1432 - }
1433 - }
1434 - }
1435 - }
1436 - }
1437 - }
1438 - }
1439 - }
1440 - }
1441 - }
1442 - }
1443 - }
1444 - }
1445 - }
1446 - }
1447 - }
1448 - }
1449 - }
1450 - }
1451 - }
1452 - }
1453 - }
1454 - }
1455 - }
1456 - }
1457 - }
1458 - }
1459 - }
1460 - }
1461 - }
1462 - }
1463 - }
1464 - }
1465 - }
1466 - }
1467 - }
1468 - }
1469 - }
1470 - }
1471 - }
1472 - }
1473 - }
1474 - }
1475 - }
1476 - }
1477 - }
1478 - }
1479 - }
1480 - }
1481 - }
1482 - }
1483 - }
1484 - }
1485 - }
1486 - }
1487 - }
1488 - }
1489 - }
1490 - }
1491 - }
1492 - }
1493 - }
1494 - }
1495 - }
1496 - }
1497 - }
1498 - }
1499 - }
1500 - }
1501 - }
1502 - }
1503 - }
1504 - }
1505 - }
1506 - }
1507 - }
1508 - }
1509 - }
1510 - }
1511 - }
1512 - }
1513 - }
1514 - }
1515 - }
1516 - }
1517 - }
1518 - }
1519 - }
1520 - }
1521 - }
1522 - }
1523 - }
1524 - }
1525 - }
1526 - }
1527 - }
1528 - }
1529 - }
1530 - }
1531 - }
1532 - }
1533 - }
1534 - }
1535 - }
1536 - }
1537 - }
1538 - }
1539 - }
1540 - }
1541 - }
1542 - }
1543 - }
1544 - }
1545 - }
1546 - }
1547 - }
1548 - }
1549 - }
1550 - }
1551 - }
1552 - }
1553 - }
1554 - }
1555 - }
1556 - }
1557 - }
1558 - }
1559 - }
1560 - }
1561 - }
1562 - }
1563 - }
1564 - }
1565 - }
1566 - }
1567 - }
1568 - }
1569 - }
1570 - }
1571 - }
1572 - }
1573 - }
1574 - }
1575 - }
1576 - }
1577 - }
1578 - }
1579 - }
1580 - }
1581 - }
1582 - }
1583 - }
1584 - }
1585 - }
1586 - }
1587 - }
1588 - }
1589 - }
1590 - }
1591 - }
1592 - }
1593 - }
1594 - }
1595 - }
1596 - }
1597 - }
1598 - }
1599 - }
1600 - }
1601 - }
1602 - }
1603 - }
1604 - }
1605 - }
1606 - }
1607 - }
1608 - }
1609 - }
1610 - }
1611 - }
1612 - }
1613 - }
1614 - }
1615 - }
1616 - }
1617 - }
1618 - }
1619 - }
1620 - }
1621 - }
1622 - }
1623 - }
1624 - }
1625 - }
1626 - }
1627 - }
1628 - }
1629 - }
1630 - }
1631 - }
1632 - }
1633 - }
1634 - }
1635 - }
1636 - }
1637 - }
1638 - }
1639 - }
1640 - }
1641 - }
1642 - }
1643 - }
1644 - }
1645 - }
1646 - }
1647 - }
1648 - }
1649 - }
1650 - }
1651 - }
1652 - }
1653 - }
1654 - }
1655 - }
1656 - }
1657 - }
1658 - }
1659 - }
1660 - }
1661 - }
1662 - }
1663 - }
1664 - }
1665 - }
1666 - }
1667 - }
1668 - }
1669 - }
1670 - }
1671 - }
1672 - }
1673 - }
1674 - }
1675 - }
1676 - }
1677 - }
1678 - }
1679 - }
1680 - }
1681 - }
1682 - }
1683 - }
1684 - }
1685 - }
1686 - }
1687 - }
1688 - }
1689 - }
1690 - }
1691 - }
1692 - }
1693 - }
1694 - }
1695 - }
1696 - }
1697 - }
1698 - }
1699 - }
1700 - }
1701 - }
1702 - }
1703 - }
1704 - }
1705 - }
1706 - }
1707 - }
1708 - }
1709 - }
1710 - }
1711 - }
1712 - }
1713 - }
1714 - }
1715 - }
1716 - }
1717 - }
1718 - }
1719 - }
1720 - }
1721 - }
1722 - }
1723 - }
1724 - }
1725 - }
1726 - }
1727 - }
1728 - }
1729 - }
1730 - }
1731 - }
1732 - }
1733 - }
1734 - }
1735 - }
1736 - }
1737 - }
1738 - }
1739 - }
1740 - }
1741 - }
1742 - }
1743 - }
1744 - }
1745 - }
1746 - }
1747 - }
1748 - }
1749 - }
1750 - }
1751 - }
1752 - }
1753 - }
1754 - }
1755 - }
1756 - }
1757 - }
1758 - }
1759 - }
1760 - }
1761 - }
1762 - }
1763 - }
1764 - }
1765 - }
1766 - }
1767 - }
1768 - }
1769 - }
1770 - }
1771 - }
1772 - }
1773 - }
1774 - }
1775 - }
1776 - }
1777 - }
1778 - }
1779 - }
1780 - }
1781 - }
1782 - }
1783 - }
1784 - }
1785 - }
1786 - }
1787 - }
1788 - }
1789 - }
1790 - }
1791 - }
1792 - }
1793 - }
1794 - }
1795 - }
1796 - }
1797 - }
1798 - }
1799 - }
1800 - }
1801 - }
1802 - }
1803 - }
1804 - }
1805 - }
1806 - }
1807 - }
1808 - }
1809 - }
1810 - }
1811 - }
1812 - }
1813 - }
1814 - }
1815 - }
1816 - }
1817 - }
1818 - }
1819 - }
1820 - }
1821 - }
1822 - }
1823 - }
1824 - }
1825 - }
1826 - }
1827 - }
1828 - }
1829 - }
1830 - }
1831 - }
1832 - }
1833 - }
1834 - }
1835 - }
1836 - }
1837 - }
1838 - }
1839 - }
1840 - }
1841 - }
1842 - }
1843 - }
1844 - }
1845 - }
1846 - }
1847 - }
1848 - }
1849 - }
1850 - }
1851 - }
1852 - }
1853 - }
1854 - }
1855 - }
1856 - }
1857 - }
1858 - }
1859 - }
1860 - }
1861 - }
1862 - }
1863 - }
1864 - }
1865 - }
1866 - }
1867 - }
1868 - }
1869 - }
1870 - }
1871 - }
1872 - }
1873 - }
1874 - }
1875 - }
1876 - }
1877 - }
1878 - }
1879 - }
1880 - }
1881 - }
1882 - }
1883 - }
1884 - }
1885 - }
1886 - }
1887 - }
1888 - }
1889 - }
1890 - }
1891 - }
1892 - }
1893 - }
1894 - }
1895 - }
1896 - }
1897 - }
1898 - }
1899 - }
1900 - }
1901 - }
1902 - }
1903 - }
1904 - }
1905 - }
1906 - }
1907 - }
1908 - }
1909 - }
1910 - }
1911 - }
1912 - }
1913 - }
1914 - }
1915 - }
1916 - }
1917 - }
1918 - }
1919 - }
1920 - }
1921 - }
1922 - }
1923 - }
1924 - }

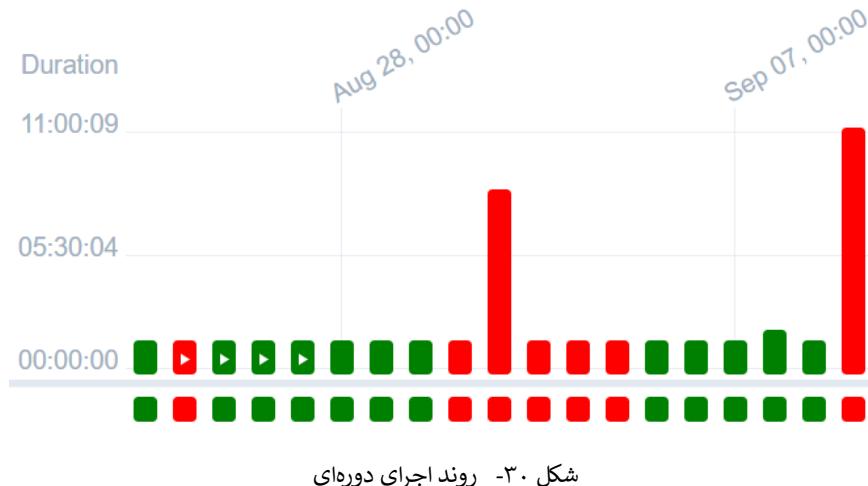
```

فصل سوم: روش پیشنهادی و نتیجه گیری

در این سناریو هر دو روش، یعنی Elasticsearch و PyBIDS، یک وظیفه‌ی یکسان را انجام داده‌اند: جستجوی فایل‌های تصویربرداری ساختاری با پسوند .nii.gz. و بازگرداندن نتایج منطبق. خروجی هر دو نیز کاملا 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 حجمی، بسیار مناسب‌تر از روش‌های مبتنی بر پیمایش مستقیم ساختار فایل‌ها مانند PyBIDS می‌سازد.

۳-۴-۳ 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 گردش‌کارهای Airflow در ایندکس‌گذاری خودکار داده‌ها پایش گردید. معیار اصلی این سناریو اطمینان از اجرای بدون خطأ، به‌روزرسانی صحیح ایندکس‌ها و حفظ یکپارچگی داده‌ها در Elasticsearch است. در ادامه به بررسی پایداری سیستم و نظم اجرای اسکریپت‌ها می‌پردازیم.



شکل ۲۰ - روند اجرای دوره‌ای

همانطور که در شکل ۳۰ مشاهده می‌شود، در این آزمایش، طی بازه‌ی زمانی مشخص تعداد ۱۹ اجرای متوالی از گردش‌کارهای تعریف شده در Airflow ثبت شد. از میان این اجراهای ۱۲ اجرای موفق و ۷ مورد ناموفق مشاهده گردید. بررسی علل خطاهای 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 سوری^{۵۳}، انتظار می‌رود تمامی اجرای‌های زمان‌بندی شده 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 اتكا است.

۵-۳ نتیجه گیری

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 تعریف شده نشان داد که اهداف پژوهه 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 طراحی شده از پویایی کافی برخوردار است و می‌تواند رشد تدریجی داده‌ها را مدیریت کند.

در سناریوی دوم، مقایسه‌ی Elasticsearch و PyBIDS 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 دقیق و صحت نتایج عملکرد یکسانی داشتند. این امر مؤید آن است که استفاده از موتور جستجوی ایندکس محور نه تنها برای کار با مجموعه‌داده‌های کوچک بلکه به‌ویژه در مواجهه با داده‌های حجمی، کارایی بالاتری خواهد داشت.

⁵² Local execution

⁵³ Stable server environment

⁵⁴ Stability

⁵⁵ Reliability

فصل سوم: روش پیشنهادی و نتیجه گیری

در سناریوی سوم، اگرچه بخشی از اجرای های دوره‌ای با خطاب 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 خطاب 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	بررسی پویایی و انعکاس تغییرات	داده‌ی جدید در تمام سطوح شناسایی و بهروز شد	معماری توانایی مدیریت رشد تدریجی داده‌ها را دارد
۲	مقایسه PyBIDS و Elasticsearch	صحت نتایج یکسان؛ سرعت ۸۳۰ برابر بیشتر	مناسب برای کاربردهای عملی و داده‌های حجمی
۳	ارزیابی پایداری اجرای خودکار	۱۲ موفق، ۷ ناموفق	روی سرور پایدار انتظار اجرای بدون خطاب 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 Airflow، Elasticsearch و Kibana) نشان می‌دهند که این رویکرد می‌تواند به عنوان یک راهکار عملی و مقیاس‌پذیر برای بایوبانک‌های داده‌های تصویربرداری مغزی مورد استفاده قرار گیرد.

۶-۳ جمع‌بندی

در این فصل ابتدا نیازمندی‌های اصلی سیستم و معماری پیشنهادی برای مدیریت و بازیابی داده‌های تصویربرداری مغزی مبتنی بر استاندارد BIDS معرفی شد. سپس مراحل مختلف پیاده‌سازی شامل استخراج و پردازش داده‌ها، ایندکس‌گذاری و ذخیره‌سازی در Elasticsearch، پایش و مصورسازی داده‌ها در Kibana، طراحی رابط کاربری برای افزودن داده‌های جدید، مأژول به‌روزرسانی خودکار با Airflow و در نهایت یکپارچه‌سازی مأژول‌ها در بستر Docker تشریح گردید.

در ادامه، روش ارزیابی بر مبنای یک مجموعه‌داده واقعی و سه سناریوی عملی شامل افزودن داده جدید، جستجو و بازیابی اطلاعات، و پایداری به‌روزرسانی‌های دوره‌ای طراحی و اجرا شد. نتایج نشان داد که سامانه‌ی پیشنهادی از نظر پویایی در مدیریت داده‌ها، سرعت و کارایی در جستجو و همچنین قابلیت اطمینان در به‌روزرسانی‌های خودکار به اهداف تعریف‌شده دست یافته است.

به طور کلی، یافته‌های این فصل بیان‌گر آن است که رویکرد ارائه‌شده توانسته است نیازمندی‌های اولیه پروژه را پوشش دهد و زیرساختی یکپارچه، مقیاس‌پذیر و قابل اتکا برای مدیریت داده‌های پیچیده تصویربرداری مغزی فراهم سازد.

مراجع

- [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, Arikan 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.

- [¹·] 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.
- [¹¹] 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>
- [¹²] 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.
- [¹³] 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
- [¹⁴] <https://bids-standard.github.io/pybids/>
- [¹⁵] <https://github.com/bids-standard/bids-validator>
- [¹⁶] 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.
- [¹⁷] 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	ایندکس‌گذاری
Retrieve	بازیابی
Biobank	بانک زیستی
Stability	پایداری
Monitoring	پایش
Data Processing	پردازش داده
Query	پرس‌و‌جو
Dynamism	پویایی
Imagery	تجسم ذهنی
Neuroimaging	تصویربرداری عصبی
Data Flow	جريان 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	مدالیته
Visualization	تصویرسازی
Scalability	مقیاس پذیری
Search Engine	موتور جستجو
Integrated	یکپارچه

پیوست‌ها و ضمایم

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 Haghghi

September 2025