

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

مبنا شهبازی

دکتر حسن حقیقی

طرح مسئله

با رشد روزافزون فناوری‌های تصویربرداری عصبی، حجم داده‌های تولیدشده به شکل چشمگیری افزایش یافته است؛ به‌طوری که برخی بایوبانک‌های بزرگ تا بیش از یک پتابایت داده نگهداری می‌کنند و پروژه‌هایی مانند Human Connectome Project به بیش از ۲۷ پتابایت داده رسیده‌اند [۱]. این داده‌ها تنها شامل تصاویر خام نیستند، بلکه شامل فراداده‌های متعدد (سن، جنسیت، شرایط اسکن، پارامترهای دستگاه و ...) نیز می‌شوند.

ویژگی مهم دیگر این داده‌ها چندوجهی بودن (Multimodality) است؛ یعنی داده‌ها از مدالیته‌های مختلفی مانند fMRI، EEG، MEG یا DTI به دست می‌آیند. هر مدالیته دیدگاهی متفاوت از مغز ارائه می‌دهد: مثلا fMRI وضوح مکانی بالا دارد در حالی که EEG وضوح زمانی بالا دارد [۲]. ترکیب این روش‌ها امکان دستیابی به تصویری جامع‌تر و دقیق‌تر از ساختار و عملکرد مغز را فراهم می‌کند [۳].

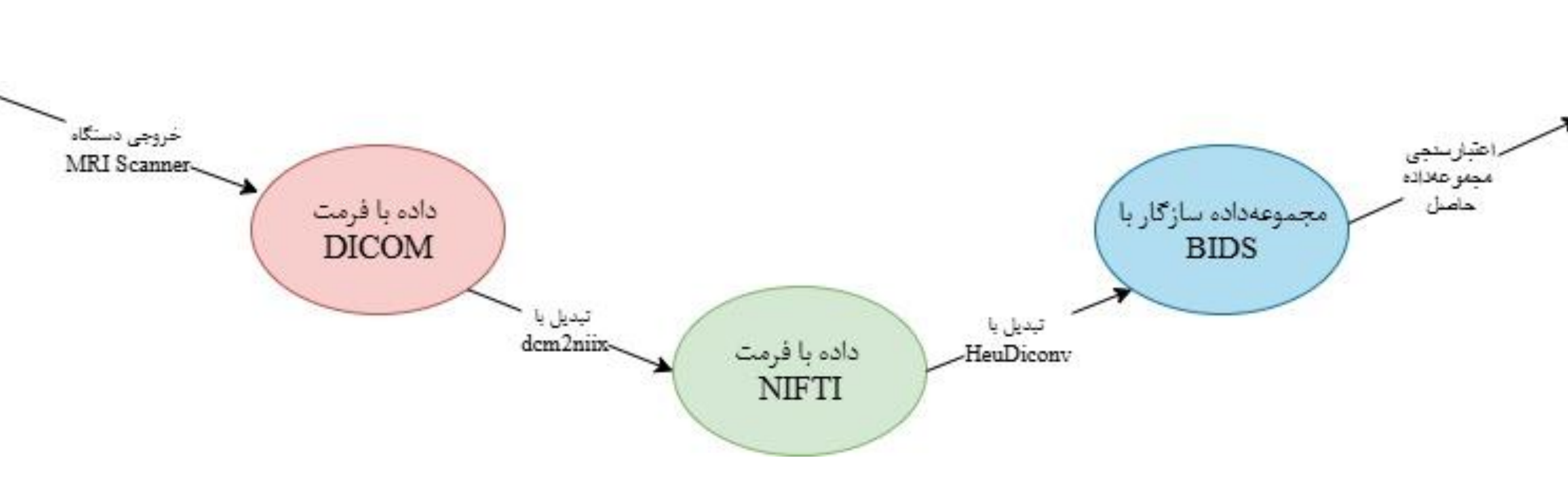
با وجود این مزایا، چندوجهی بودن داده‌ها چالش‌های جدی در ذخیره‌سازی، جستوجو و بازیابی سریع به همراه دارد. بنابراین، نیاز به یک زیرساخت یکپارچه و کارآمد برای مدیریت داده‌ها وجود دارد؛ زیرساختی که بتواند هم داده‌ها را به‌صورت سازمان‌یافته ذخیره کند، هم جست‌وجوی سریع و تحلیل بلادرنگ را امکان‌پذیر سازد.

ادبیات پژوهش

تصویربرداری مغزی در طول زمان از نظر قالب‌های ذخیره‌سازی دچار تحول شده است. در ابتدا، فرمت DICOM به‌عنوان خروجی اصلی دستگاه‌های تصویربرداری مورد استفاده قرار می‌گرفت که تصویر را همراه با فراداده‌های بالینی نگهداری می‌کرد. هرچند این فرمت برای تبادل و بایگانی مناسب بود، اما برای پردازش‌های محاسباتی کارآمد نبود [۴].

به همین دلیل، فرمت NIfTI معرفی شد که ساختاری ساده‌تر و مناسب برای پردازش داده‌های سه‌بعدی و چهاربعدی تصویربرداری فراهم می‌کرد [۵]. گام مهم بعدی معرفی استاندارد BIDS بود که نه تنها داده‌ها را به‌صورت منظم و سلسله‌مراتبی سازمان‌دهی می‌کرد، بلکه فراداده‌ها و فایل‌های متنی مکمل (JSON و TSV) را نیز در ساختار خود جای داد. این استاندارد، امکان اشتراک‌گذاری آسان و اجرای خطوط پردازش خودکار را فراهم نمود و به سرعت به یکی از پرکاربردترین قالب‌ها در علوم اعصاب محاسباتی بدل شد [۶].

در کنار BIDS، ابزارهای متعددی برای تسهیل تبدیل و سازمان‌دهی داده‌ها توسعه یافته‌اند. ابزار dcm2nii از داده‌های خام DICOM را به فرمت NIfTI و فایل‌های جانبی JSON تبدیل می‌کند و بخش مهمی از ناسازگاری‌های ناشی از تفاوت پیاده‌سازی دستگاه‌های مختلف را برطرف می‌سازد [۷]. HeuDiConv با استفاده از قواعد ابتکاری، امکان سازمان‌دهی انعطاف‌پذیر و خودکار پوشه‌ها و فایل‌ها را فراهم می‌کند و به همین دلیل در پروژه‌های بزرگ و بایوبانک‌ها کاربرد گسترده‌ای دارد [۸]. همچنین، BIDS Validator ابزار مهمی برای کنترل کیفیت و ارزیابی انطباق داده‌ها با استاندارد BIDS است که از خطاهای ساختاری و متادیتایی جلوگیری می‌کند. این مسیر تحولی را می‌توان در شکل ۱ مشاهده کرد.



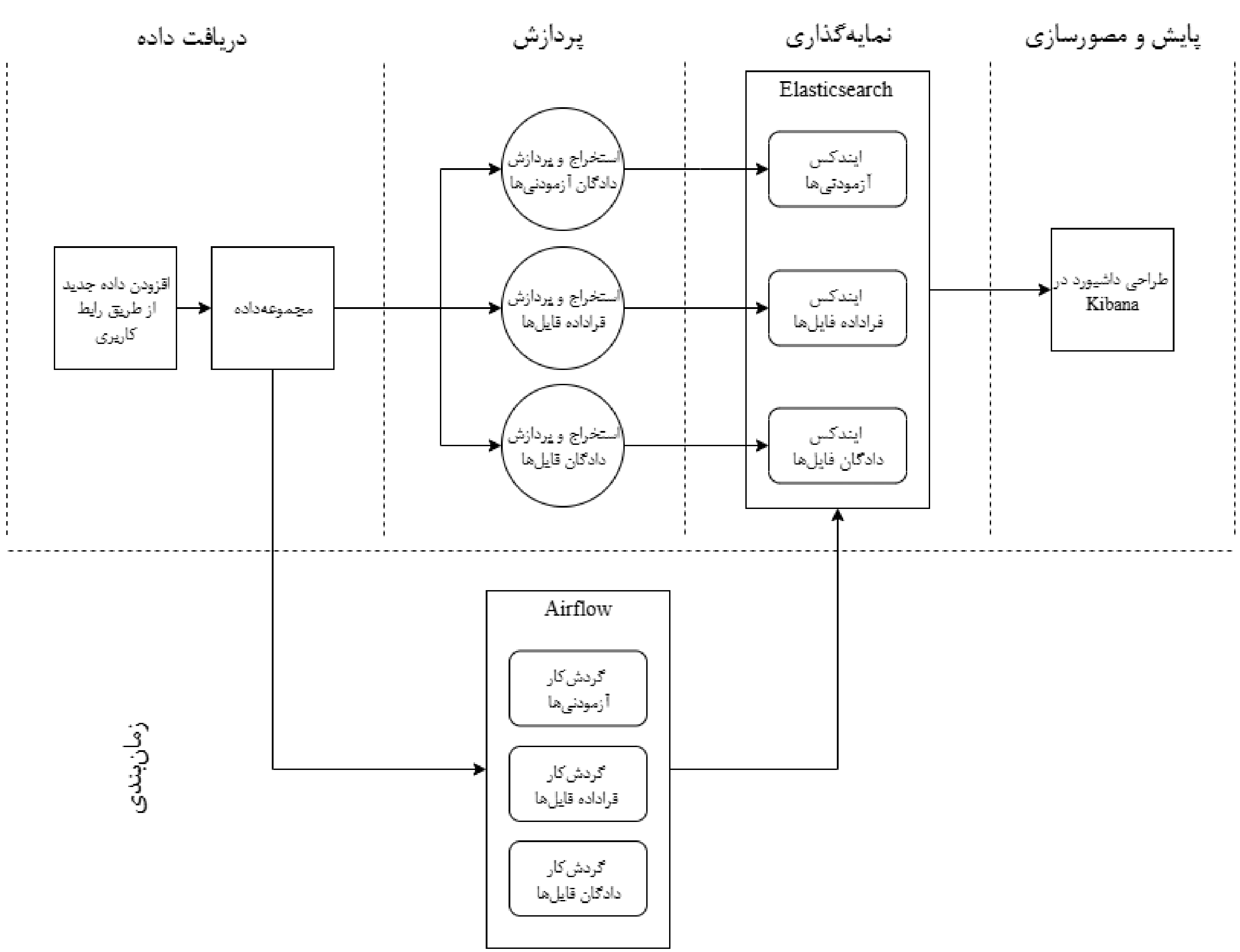
شکل ۱- فرایند تبدیل داده خام به قالب BIDS

این ابزارها در کنار هم، یک زنجیره کامل برای آماده‌سازی و مدیریت داده‌های تصویربرداری فراهم می‌کنند: از تبدیل داده خام تا سازمان‌دهی و اعتبارسنجی. به‌ویژه در بایوبانک‌های بزرگ که داده‌ها به‌صورت مستمر و حجیم تولید می‌شوند، وجود چنین اکوسیستمی اهمیت ویژه‌ای دارد، زیرا بدون آن، یکپارچگی داده‌ها از بین رفته و امکان تحلیل‌های خودکار و اشتراک‌گذاری داده‌ها به شدت محدود خواهد شد.

مرور کارهای انجام‌شده

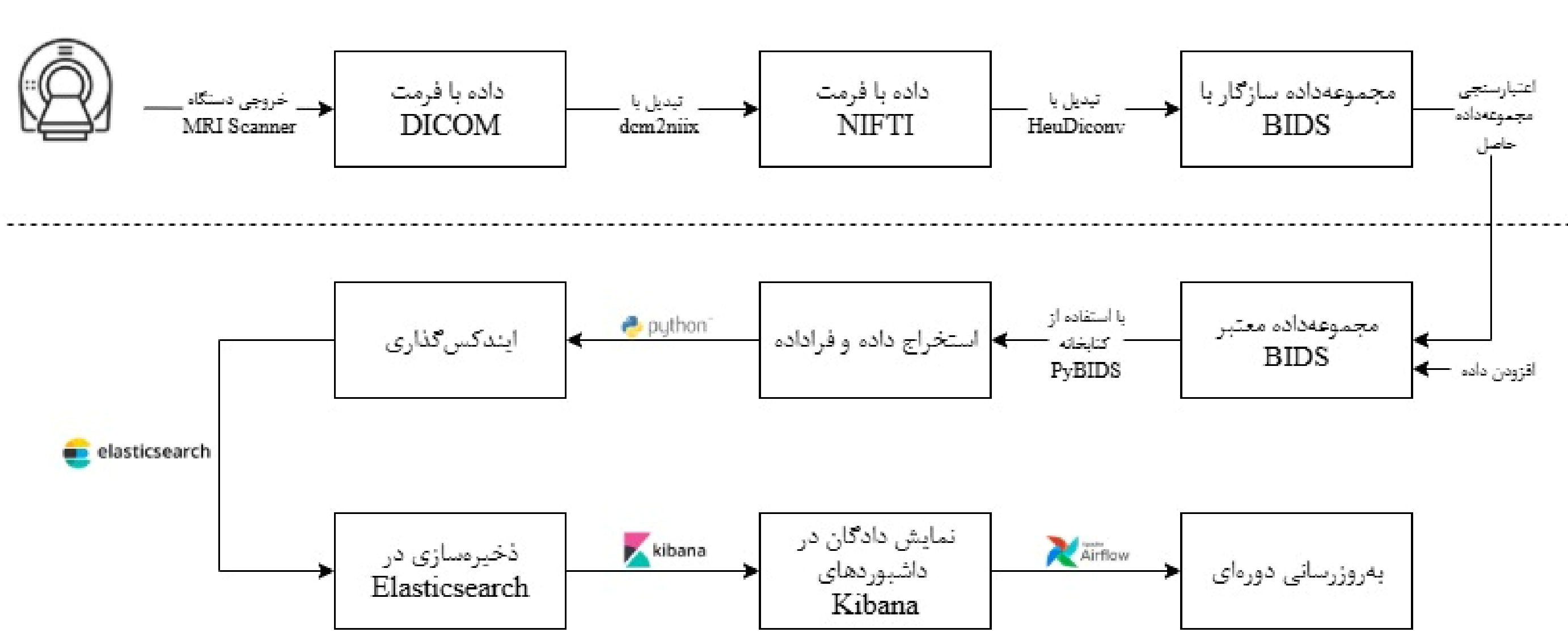
هدف اصلی این پروژه، طراحی و پیاده‌سازی یک زیرساخت مدیریت داده برای بایوبانک‌های تصویربرداری مغزی بر پایه استاندارد BIDS بود. این زیرساخت باید امکان ذخیره‌سازی منظم داده‌ها، جستجوی سریع و انعطاف‌پذیر، افزودن آزمودنی‌های جدید از طریق یک رابط کاربری تحت وب، و نیز پایش و تحلیل اطلاعات از طریق داشبوردهای تصویری را فراهم سازد. علاوه بر این، سیستم باید از پایداری و مقیاس‌پذیری کافی برای مواجهه با حجم رو به رشد داده‌های چندوجهی برخوردار باشد.

معماری پیشنهادی برای دستیابی به این اهداف در پنج لایه‌ی اصلی طراحی شد (شکل ۲). در لایه نخست، داده‌های جدید از طریق رابط کاربری Flask دریافت و در قالب استاندارد BIDS سازمان‌دهی می‌شوند. سپس در لایه پردازش، اسکریپت‌های پایتونی داده‌ها و فراداده‌های موجود در فایل‌های تصویری، JSON و TSV را استخراج و آماده‌سازی می‌کنند. این اطلاعات در لایه سوم به سامانه‌ی Elasticsearch منتقل و در قالب ایندکس‌های مجزا ذخیره می‌شوند تا امکان جستجو و بازیابی سریع داده‌ها فراهم شود. در گام بعد، لایه زمان‌بندی و به‌روزرسانی با استفاده از Airflow اجرای خودکار پردازش‌ها و ایندکس‌گذاری را تضمین می‌کند. در نهایت، لایه پایش و مصورسازی با بهره‌گیری از Kibana داشبوردهایی فراهم می‌سازد که داده‌ها را به صورت بصری و تحلیلی در اختیار پژوهشگران قرار می‌دهند.



شکل ۲- نمای معماری کلان سیستم

جریان داده‌ها از لحظه ورود تا مصورسازی نیز در نمودار شکل ۳ نشان داده شده است که فرایند پس از خط‌چین در محدوده این پروژه جای می‌گیرد. این معماری، هسته اصلی سیستم پیشنهادی را شکل داده و ارتباط اجزا را به‌روشنی ترسیم می‌کند.



شکل ۳- جریان دادگان از تصویربرداری تا مصورسازی

نتیجه‌گیری و روال‌های موجود برای کارهای آتی

برای ارزیابی عملکرد روش پیشنهادی، مجموعه‌داده OpenNeuro ds006040 شامل داده‌های fMRI، DWI و EEG از ۲۸ آزمودنی استفاده شد [۹]. سه سناریوی کلیدی در ارزیابی تعریف گردید:

- افزودن داده جدید:** تایید شد که سیستم به‌طور پویا تغییرات را در ایندکس‌ها و داشبوردها منعکس می‌کند.
- جستجو و بازیابی:** مقایسه کتابخانه پایتونی PyBIDS و Elasticsearch نشان داد که در پاسخ به یک پرس‌وجوی یکسان، از نظر صحت هر دو نتایج مشابهی داشتند اما Elasticsearch حدود ۸۳۰ برابر سریع‌تر عمل کرد. جزئیات نتایج این سناریو در جدول ۱ قابل مشاهده است.

جدول ۱- مقایسه نتایج دو روش جستجو باPyBIDS و Elasticsearch		
PyBIDS	Elasticsearch	
۲۸	۲۸	تعداد سند یافت شده
۶.۶۵۱۳	۰.۰۰۸	زمان اجرا (ثانیه)

- به‌روزرسانی دوره‌ای:** در اجرای‌های محلی بخشی از وظایف با خطا مواجه شد اما مشخص شد که معماری پیشنهادی در یک محیط پایدار می‌تواند به‌طور مطمئن عمل کند.

در مجموع، سیستم پیشنهادی توانست نیازهای اساسی موردنظر را پوشش دهد و نشان داد که رویکرد ایندکس‌محور همراه با ابزارهای متن‌باز، یک راهکار عملی و مقیاس‌پذیر برای مدیریت داده‌های تصویربرداری مغزی است. خلاصه ای از نتایج ارزیابی در جدول ۲ قابل مشاهده است. با وجود این دستاوردها، مسیرهای بسیاری برای توسعه آتی باقی است؛ از جمله گسترش سیستم برای پشتیبانی از داده‌های حجیم‌تر، افزودن قابلیت تحلیل‌های پیشرفته در Kibana و یکپارچه‌سازی با الگوریتم‌های یادگیری ماشین برای پردازش خودکار داده‌های چندوجهی.

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 Elasticsearch 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

منابع

[۱] 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.

[۲] 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>

[۳] 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.

[۴] 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.

[۵] Clunie, D. A. (2000). DICOM Structured Reporting. PixelMed Publishing.

[۶] 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.

[۷] 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>

[۸] 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

[۹] <https://openneuro.org/>