

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

مبینا شهبازی

دکتر حسن حقیقی

نتیجه‌گیری و روال‌های موجود برای کارهای آتی

برای ارزیابی عملکرد روش پیشنهادی، مجموعه داده OpenNeuro ds006040 شامل داده‌های EEG و DWI از استاندارد BIDS بود. این زیرساخت باید امکان ذخیره‌سازی منظم داده‌ها، جستجوی سریع و انعطاف‌پذیر، افزوندن آزمودنی‌های جدید از طریق یک رابط کاربری تخت و ب، 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 بتواند عملکرد خوبی داشته باشد.

- آزمودنی‌های جدید: تایید شد که سیستم به طور پویا تغییرات را در ایندکس‌ها و داشبوردهای منعکس می‌کند.
- جستجو و بازیابی: مقایسه کتابخانه پایتونی PyBIDS و مقیاس‌پذیری کافی برای مواجهه با حجم رو به رشد داده‌های 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 با PyBIDS و Elasticsearch

Elasticsearch	PyBIDS
۲۸	۲۸
تعداد سند یافت شده	۶.۶۵۱۳
زمان اجرا (ثانیه)	۰.۰۰۸

- پژوهشی دوره‌ای: در اجرای‌های محلی بخشی از وظایف با خطأ مواجه شد اما مشخص شد که معماری پیشنهادی در یک محیط پایدار می‌تواند به طور مطمئن عمل کند.

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

جدول ۲- نتایج ارزیابی

تفصیل	سناریو	هدف ارزیابی	نتیجه ارزیابی
معماری توانایی مدیریت رشد	۱	داده‌ی جدید در تمام سطوح بررسی یونایی و انعکاس شناسایی و بروز شد تغییرات	افزوند داده جدید
مناسب برای کاربردهای عملی و داده‌های جی‌جی	۲	صحت نتایج یکسان؛ سرعت مقایسه PyBIDS و Elasticsearch و جستجو و بازیابی	۸۳۰ برابر بریزتر
روی سرور پایدار انتظار اجرای بدون خطا وجود دارد	۳	به روزرسانی دوره‌ای ارزیابی پایداری اجرای خودکار ۷ ناموفق	۱۲ موفق

منابع

- [1] 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.
- [2] 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>
- [3] Tulay EE, Metin B, Tarhan N, Arican MK. Multimodal Neuroimaging: Basic Concepts and Classification of Neuropsychiatric Diseases. Clin EEG Neurosci. 2019 Jan;50(1):20-33. doi: 10.1177/155059418782093. Epub 2018 Jun 20. PMID: 29925268.
- [4] 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.
- [5] Clunie, D. A. (2000). DICOM Structured Reporting. PixelMed Publishing.
- [6] Cox, R.W. & Ashburner, John & Breman, Hester & Fissell, Kate & Haselgrave, 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.
- [7] Yaroslav O. Halchenko, Mathias Goncalves, Sarajit Ghosh, Pablo Velasco, Matteo Visconti di Oleggio Castello, Taylor Salo, John T. Wodder II, Michael Hanke, Patrick Sadik, Krzysztof Jacek Gorgolewski, Horea-Ioan Ioanias, 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>
- [8] 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
- [9] <https://openneuro.org/>

مرور کارهای انجام‌شده

هدف اصلی این پروژه، طراحی و پیاده‌سازی یک زیرساخت مدیریت داده برای بایوانک‌های تصویربرداری مغزی برایه است. بطوری که برخی بایوانک‌های بزرگ تا بیش از یک پتاپایت داده نگهداری می‌کنند و پروژه‌هایی مانند Human Connectome Project به بیش از ۲۷ پتاپایت 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 مدل‌الیته‌های مختلفی مانند EEG، fMRI، MEG یا DTI به دست می‌آیند. هر مدل‌الیته دیدگاهی مختلف از مغز ارائه می‌دهد: مثلاً fMRI وضوح زمانی بالا دارد در حالی که EEG وظایف ایجاد می‌کند. 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، جستجو 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 بصورت سازمان‌یافته ذخیره کند، هم جستجوی سریع و تحلیل بلادرنگ را امکان‌پذیر سازد.

طرح مسئله

با رشد روزافزون فناوری‌های تصویربرداری عصبی، حجم داده‌های تولید شده به شکل چشمگیری افزایش یافته است؛

Human Connectome Project به بیش از ۲۷ پتاپایت داده رسیده‌اند [۱]. این داده‌ها تنها شامل تصاویر خام نیستند، بلکه

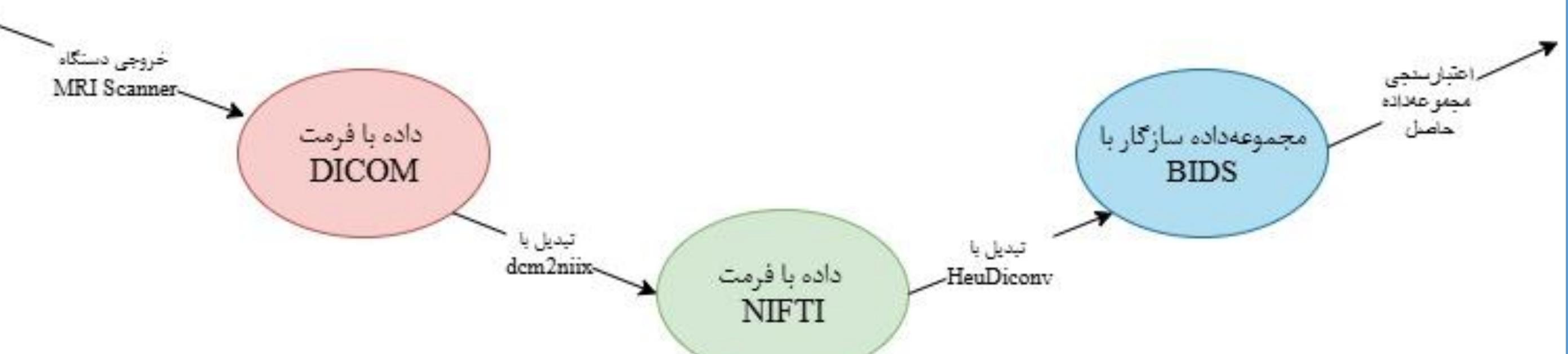
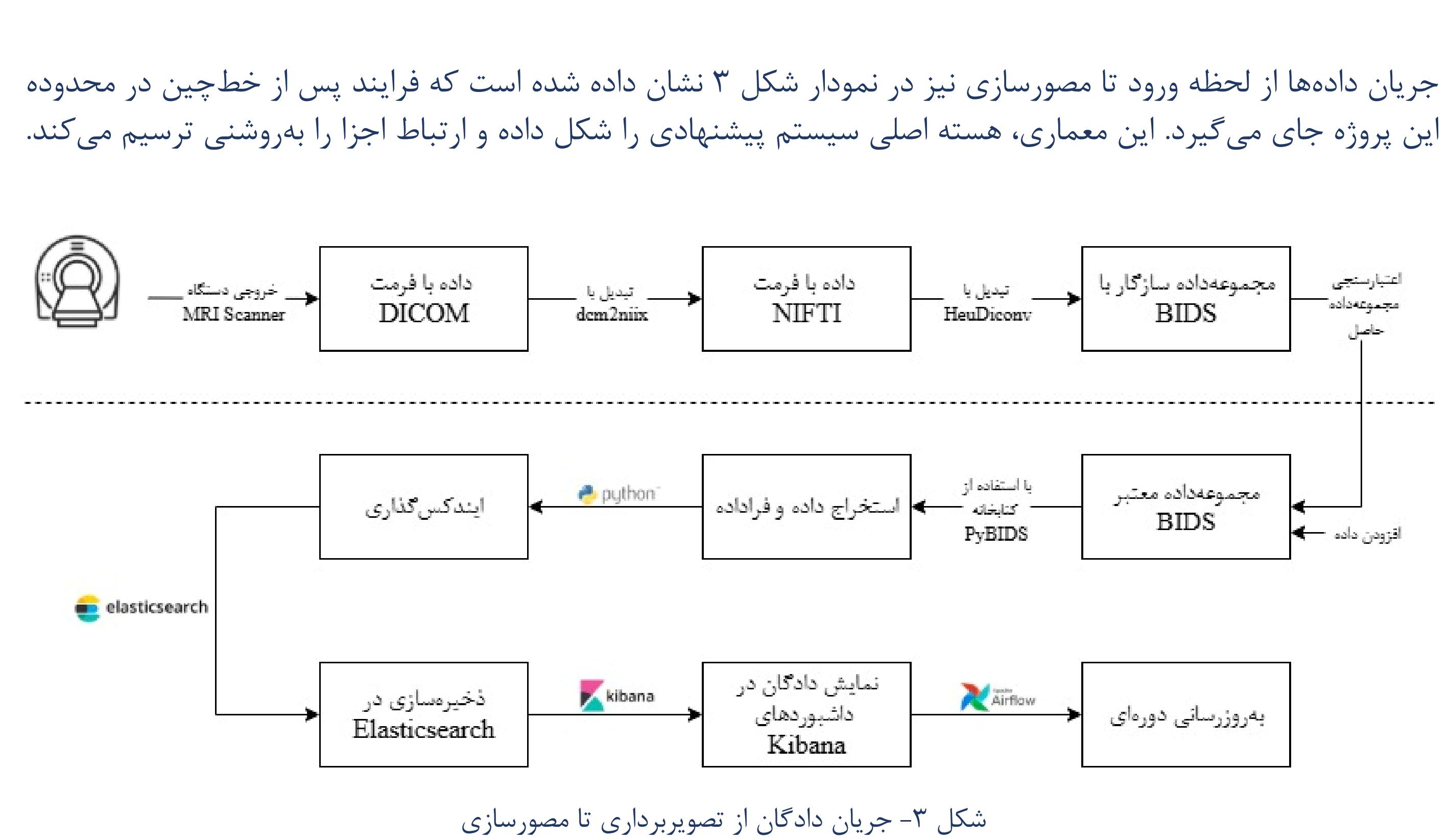
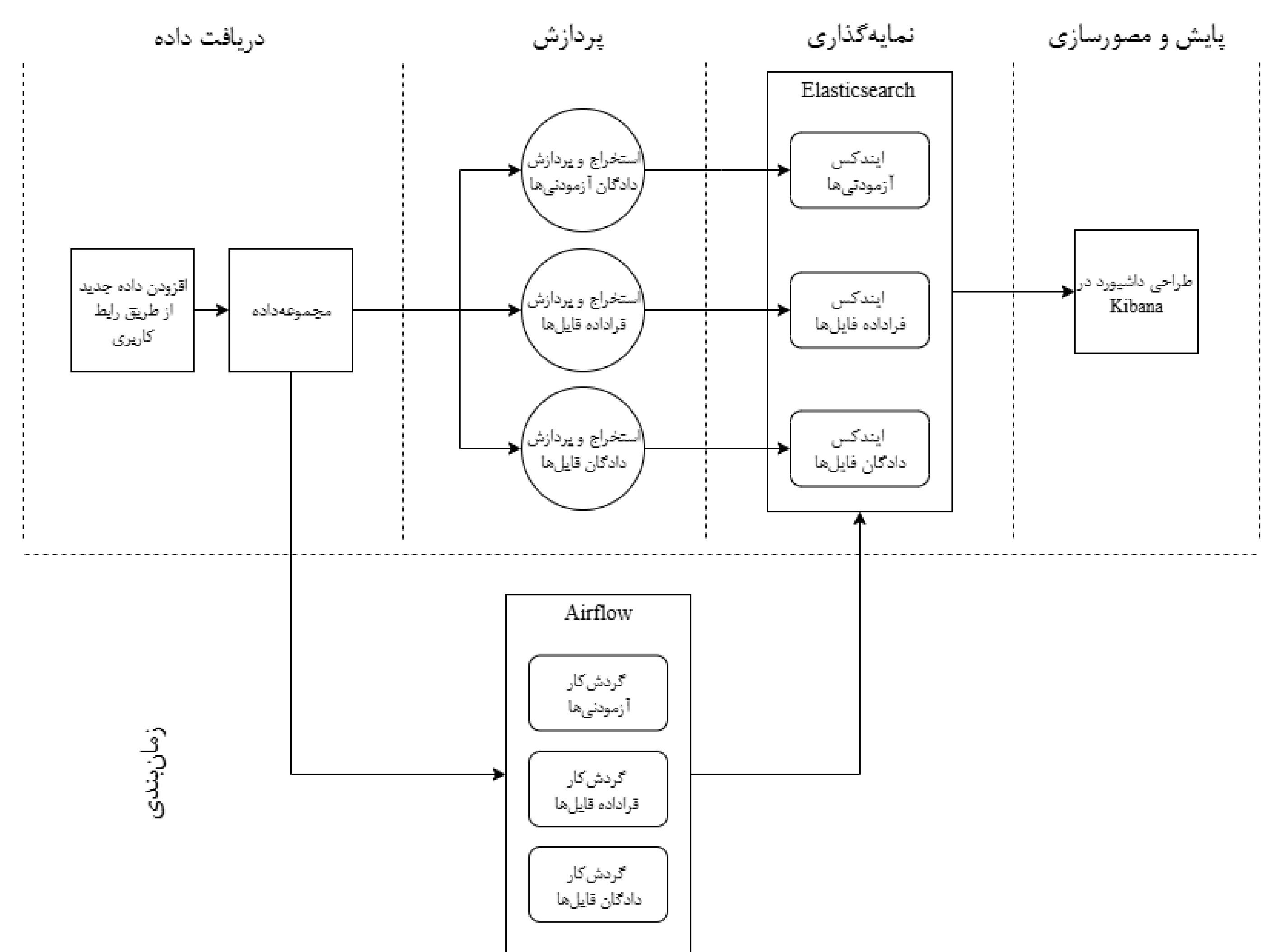
ویژگی مهم دیگر این داده‌ها چندوجهی بودن (Multimodality) است؛ یعنی داده‌ها از مدل‌الیته‌های مختلفی مانند Elasticsearch، Flask دریافت و در قالب استاندارد BIDS سازمان‌دهی می‌شوند. سپس در لایه پردازش، Elasticsearch و اسکریپت‌های پایتونی داده‌ها و فراداده‌های موجود در فایل‌های تصویری، TSV و JSON را استخراج و آماده‌سازی می‌کنند. این اطلاعات در لایه سوم به سامانه Elasticsearch منتقل و در قالب ایندکس‌های مجزا ذخیره می‌شوند تا امکان جستجو و بازیابی سریع داده‌ها وجود شود. در گام بعد، لایه زمان‌بندی و بروزرسانی 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.

ادیبات پژوهش

تصویربرداری مغزی در طول زمان از نظر قالب‌های ذخیره‌سازی دچار تحول شده است. در ابتدا، فرمت DICOM به عنوان خروجی اصلی دستگاه‌های تصویربرداری مورد استفاده قرار می‌گرفت که تصویر را همراه با پایگان 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 داده‌ها رسیدگی و چهار بعدی تصویربرداری سازمان‌دهی می‌کرد، بلکه گام بعدی معرفی استاندارد BIDS بود که نه تنها داده‌ها را به صورت منظم و سلسه‌مرتبی سازمان‌دهی می‌کرد، بلکه فراداده‌ها و خودکار پوشش‌ها و فایل‌ها را فراهم می‌کرد و همین دلیل در پژوهش‌های سه بعدی و چهار بعدی این داده‌ها را به صورت خودکار از جای خود جدا نمودند [۵]. همچنان، Airflow این داده‌ها را به صورت منظم و مکمل (TSV و JSON) 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 پکاربریدترین قالب‌ها در علوم اعصاب محاسباتی بدل شد [۶].

در کار BIDS، ابزارهای متعددی برای تسهیل تبدیل و سازمان‌دهی داده‌ها توسعه یافته‌اند. در ابتدا، فرمت dcm2niiix ابزاری است که 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 اینتکاری، امکان سازمان‌دهی انعطاف‌پذیر و خودکار پوشش‌ها و فایل‌ها را فراهم می‌کند و به همین دلیل در پژوهش‌های بزرگ و بایوانک‌ها کاربرد گسترهای دارد [۸]. همچنین، BIDS Validator ابزار مهمی برای کنترل کیفیت و ارزیابی انتظام داده‌ها با استاندارد BIDS است که از خطاهای ساختاری و متاداده‌ای جلوگیری می‌کند. این مسیر تحولی را می‌توان در شکل ۱ مشاهده کرد.



این ابزارها در کنار هم، یک زنجیره کامل برای آماده‌سازی و مدیریت داده‌های تصویربرداری فراهم می‌کنند: از تبدیل داده خام تا سازمان‌دهی و اعتبارسنجی. بعویه در بایوانک‌های بزرگ که داده‌ها به صورت مستمر و حجمی 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.