An Agentic AI Framework for Unified Brain Health Profiling: Integrating Multimodal Neurological Data with Foundation Models

Anonymous Author(s)
Affiliation
Address
`email`

Abstract

The diagnosis and monitoring of neurological health are hindered by the fragmentation of clinical data, where diagnostic tests such as Quantitative Electroencephalography (QEEG), Videonystagmography (VNG), and imaging studies operate in isolation. This paper introduces the Brain Profiling System, an agentic AI framework designed to address this challenge by integrating multimodal neurological data into a unified, clinically coherent profile. Leveraging foundation models, our system comprises three core modules: **MedRecs**, a document intelligence engine using Retrieval and Cache Augmented Generation (RAG/CAG) Lewis et al. [2020] to summarize unstructured medical records with full citation traceability; **MedSight**, a multimodal AI for analyzing medical images and diagnostic graphs; and **Clinical Synthesis**, an engine that generates standardized SOAP notes from all integrated data sources. Our evaluation demonstrates transformative efficiency gains, reducing medical record review time from over two days to a target of three hours and complex imaging analysis from over 1.5 hours to under a minute. This framework not only streamlines clinical workflows but also enhances the defensibility of medical documentation for legal and insurance applications, establishing a scalable platform for future neurological research.

1 Introduction

The assessment of neurological health, particularly following traumatic brain injuries (TBIs), involves a diverse array of diagnostic modalities, including CNS Vital Signs (CNSVS), QEEG, Posturography, and medical imaging. While each test provides a critical view of brain function, the data remains siloed, forcing clinicians, legal professionals, and insurers to manually synthesize information from disparate and often unstructured sources. This fragmentation leads to diagnostic delays, increased administrative burden, and potential for missed correlations between different neural systems.

To address these inefficiencies, we have developed the Brain Profiling System, a comprehensive platform that unifies fragmented neurological data into a single, actionable health profile. The system is built upon an agentic AI architecture that leverages foundation models to automate the extraction, analysis, and synthesis of complex medical information Wang et al. [2025]. It is composed of three primary modules: **MedRecs** for intelligent summarization of medical records, **MedSight** for the interpretation of visual diagnostics, and **Clinical Synthesis** for generating holistic, standardized clinical reports. This paper details the system's architecture, methodology, and performance, demonstrating its capacity to dramatically reduce clinical workloads while improving the accuracy and transparency of brain health assessment.

Submitted to 2nd Conference on Applied AI and Scientific Machine Learning (CASML 2025). Do not distribute.

2 System Architecture and Methodology

The Brain Profiling System is a modular, HIPAA-compliant platform designed to process raw, multimodal inputs and produce standardized, interpretable outputs. Its architecture integrates several specialized AI components, each targeting a specific challenge in the clinical workflow.

2.1 MedRecs: Document Intelligence Engine

MedRecs serves as the system's data ingestion and summarization backbone, designed to process hundreds of pages of disorganized medical records. It employs a sophisticated pipeline combining Optical Character Recognition (OCR), Non Optical Character Recognition (Non-OCR) and Natural Language Processing (NLP) to convert scanned documents, lab reports, and handwritten notes into a structured, chronological timeline with citations.

The core of MedRecs is a Retrieval and Cache-Augmented Generation system Gao et al. [2023], built using foundation model of MedGemma. As illustrated in Figure 1, documents are ingested, chunked, embedded, and stored in a vector databases. User queries trigger a retrieval process that injects relevant context into a model to generate cited, accurate answers, eliminating hallucinations Asai et al. [2023]. A specialized **ResearChat** agent, trained on over 4,800 brain injury research papers and synced with PubMed and ArXiv, provides real-time, evidence-based context to clinical findings, bridging the gap between patient data and scientific literature.

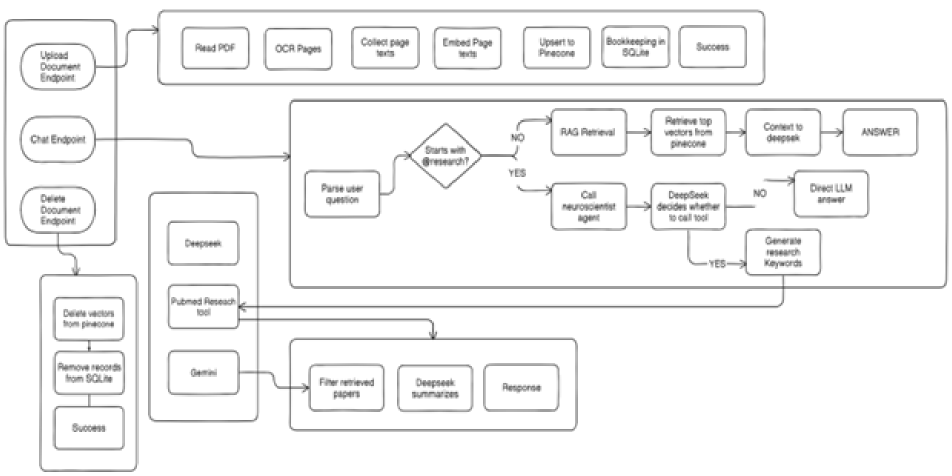

Figure 1: System architecture diagram for the MedRecs module, illustrating the data flow for document ingestion, deletion, and the RAG-based chat functionality with a specialized research agent.

2.2 MedSight: Multimodal Diagnostic Analysis

MedSight addresses the interpretation of visual medical data, a significant bottleneck in clinical practice. It leverages multimodal foundation model of MedGemma to analyze a wide range of inputs, including MRI scans, CT Scans, EEG plots, posturography graphs, and VNG eye movement tests. MedSight generates both highly technical summaries for clinicians and clear, patient-friendly explanations, demystifying complex results for all stakeholders. Every finding can be cross-referenced with current medical literature to ensure credibility and traceability.

2.3 Clinical Synthesis Engine

The Clinical Synthesis engine is the final integration point where all processed data converges Our custom-built dart-boxes have an algorithm set which triggers an agentic run on disparate PDF data of CNS Vital Signs (CNSVS), QEEG, Posturography, Electroencephalography (QEEG), Videonystagmography (VNG) and runs on low context parameters making it highly capable to further calculations and interpretations while keeping it prone from standard large-language model hallucinations. It ingests the structured outputs from MedRecs and MedSight, along with data from four digital patient questionnaires. This consolidated information is transformed into standardized SOAP (Subjective, Objective, Assessment, Plan) notes, which include diagnostic categorizations and draft treatment plans complete with rationales and literature citations. This gives clinicians critical time over clerical tasks and it incorporates a continuous learning loop, refining its models as more patient data is processed to improve diagnostic precision over time.

3 Results and Evaluation

The implementation of the Brain Profiling System has yielded dramatic improvements in efficiency across a range of diagnostic tasks. The system significantly reduces the time required for data handling and interpretation, freeing clinicians to focus on direct patient care. Table 1 summarizes the performance gains observed with the system's modules.

The most notable gains are seen in tasks that are traditionally time-intensive. For instance, a comprehensive review of a 1,000-page medical record, which previously took a clinician approximately two full workdays, can now be completed in under four hours with the MedRecs module. Similarly, MedSight reduced the time for radiology imaging analysis from over three hours to under three minutes—a greater than 99% reduction.

Table 1: Performance Gains in Diagnostic Task Completion Times

Diagnostic Task	Time Before	Time After	Reduction
Cognitive & Limbic Vitals	45 mins	3 mins	93%
EEG Calculations & Interpretation	3 hours	4 mins	95%
Videonystagmography Interpretation	60 mins	2 mins	93%
1000-page Medical Records Review	2 days	< 4 hours	>90%
Radiology Imaging Analysis	3+ hours	3 mins	>99%

Beyond quantitative metrics, the system enables enhanced qualitative insights. By correlating data across different modalities—such as linking QEEG asymmetries with gait instability observed in posturography—clinicians can identify complex patterns that were previously obscured by data silos. This holistic view is particularly valuable in diagnosing and managing TBIs, where injuries often affect multiple neurological systems.

4 Discussion and Impact

The Brain Profiling System represents a paradigm shift from fragmented, manual data analysis to integrated, AI-driven neurological assessment. Its impact extends across clinical, legal, and research domains.

In **clinical practice**, the system acts as an assistive tool that augments, rather than replaces, human expertise. By automating tedious documentation and providing evidence-based insights at the point of care, it allows clinicians to make faster, more accurate, and legally defensible decisions.

For the **legal and insurance sectors**, the platform provides standardized, source-cited reports that distill thousands of pages of medical records into clear, defensible narratives. This drastically reduces the "time-to-comprehension" for attorneys and enables insurers to more accurately quantify neurological impairment, potentially through a data-driven "brain health score".

For **research**, the system creates a scalable infrastructure for large-scale observational studies. By aggregating anonymized, multimodal data, it builds a living dataset that can be used to define new normative baselines, test novel algorithms, and uncover previously hidden correlations in brain injury and recovery mechanisms, thereby accelerating discovery in neuroscience and precision medicine.

5 Conclusion

The Brain Profiling System successfully demonstrates the application of foundation models within an agentic AI framework to address the long-standing challenge of data fragmentation in neurological healthcare. By integrating multimodal diagnostics into a single, coherent profile, our system delivers significant efficiency gains, enhances diagnostic accuracy, and provides robust, defensible documentation for clinical and legal use. It transforms scattered data into structured intelligence, empowering professionals across medicine, law, and research to make more informed decisions. This work lays the foundation for a future where brain health is understood not as a collection of isolated data points, but as a holistic, dynamic, and fully integrated picture.

References

References

Akari Asai, Sida Lin, and Danqi Chen. Self-rag: Learning to retrieve, generate, and critique through self-reflection. *arXiv preprint arXiv:2310.11511*, 2023. doi: 10.48550/arXiv.2310.11511.

Yunfan Gao, Ziyi Wang, and Haofen Chen. Retrieval-augmented generation for large language models: A survey. *arXiv preprint arXiv:2312.10997*, 2023. doi: 10.48550/arXiv.2312.10997.

Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Myle Ott, Wen-tau Chen, Alexis Conneau, et al. Retrieval-augmented generation for knowledge-intensive nlp tasks. *arXiv preprint arXiv:2005.11401*, 2020. doi: 10.48550/arXiv.2005.11401.

Xinyu Wang, Zhen Wu, and Hai Zhao. Octotools: An agentic framework with extensible tools for complex reasoning. *arXiv preprint arXiv:2502.11271*, 2025. doi: 10.48550/arXiv.2502.11271.