of 0.684 and good calibration along the identity line up to predicted probability of 0.3. DISCUSSION/SIGNIFICANCE OF IMPACT: A LRM to predict rEF using an ECG is computationally efficient and address important barriers of adoption in the EMS setting. It does not require additional EMS training or equipment. LRM transparency facilitates communication between paramedics and medical control clinicians to make shared decisions regarding treatment and transport decisions.

54

Broad and shallow, deep and narrow, data track health's causal arrow

Jun Ha Chang[1], Ran Dai[1], Anuj Sharma[2], Soumik Sarkar[2], Elizabeth Vlock[1], Carrie Bailey[1], Laura Lienemann[1] and Matthew Rizzo[3]
[1]University of Nebraska Medical Center; [2]Iowa State University and [3]University of Nebraska

OBJECTIVES/GOALS: Broad and shallow data from large national cohorts provide generalizable real-world insights, while deep and narrow data from smaller cohorts capture detailed multimodal measures. This project integrates both approaches to study cognitive decline in Parkinson's disease, which affects up to 50% of patients. METHODS/STUDY POPULATION: This longitudinal study includes validated digital biomarkers of motor, non-motor, sleep, and driving activity in 150 participants, plus novel social connectedness measures in 150 PD – partner dyads. PD and cognitive status are assessed annually; 4 weeks of continuous real-world behavior via actigraphy, social connectedness, and driving sensors are collected biannually. Anchor variables from pilot PD, PPMI, and LongROAD enable harmonization and cross-cohort modeling. Model development focuses on generative time series models that can capture the joint distribution of multimodal time series data for predicting values for future time steps. Together combined with genetics, plasma, and neuroimaging, they enable digital twins to predict decline and identify modifiable risks. RESULTS/ANTICIPATED RESULTS: This project is designed to make impactful contributions to scientific knowledge, technical capability, clinical practice, and healthcare equity. Addressing key gaps in understanding PD cognitive decline, developing clinically validated digital biomarkers, and advancing data modeling for public health and precision medicine, including the creation of Digital Twin models. By integrating real-world data and social context into PD research, we will enhance clinical trials, improve patient outcomes, and provide clinicians with valuable tools, leading to a deeper understanding of PD and ADRDs. The data generated will be mined and shared with colleagues for years, driving new research and insights. DISCUSSION/SIGNIFICANCE OF IMPACT: This project enhances monitoring and prediction of cognitive decline in PD, developing clinically valid, interpretable tools for real-world health. This generalizable platform unifies digital, imaging, and biomarker data as a model for translational research, advancing trial readiness across neurological and chronic diseases.

55

Research intelligence for women's health: Preliminary results from an Ask-Your-Data Utility powered by large language models

Jungwei Fan and Sivaraman Rajaganapathy
Mayo Clinic

OBJECTIVES/GOALS: To contextualize existing knowledge in translational science, women's health research can benefit from artificial intelligence (AI) methods for obtaining insights out of massive scientific literature. This study explored the feasibility of making an AI model answer diverse questions based on a set of our institution-specific publications. METHODS/STUDY POPULATION: We developed an automated pipeline to retrieve PubMed publications on women's health that acknowledged the Mayo Clinic Clinical and Translational Science Award of the past two cycles. The publications need to be indexed with MeSH terms (as a major topic) including "Women's Health", "Obstetrics", "Gynecology", "Female Urogenital Diseases and Pregnancy Complications", and "Perinatal Care", etc. In addition, we enforced the species to be "Human" and the demographics to include "Female." The metadata of every article including title, abstract, publication year, journal, and author affiliation were saved into a single JSON file. The JSON file was then used as the input context along with a few test prompts submitted to the gemini-2.5-flash model. All the experiments were conducted in a Google Colab environment. RESULTS/ANTICIPATED RESULTS: A total of 237 articles on women's health were retrieved for the experiment. The AI model correctly summarized the top 5 journals but did not obtain the correct article counts until explicitly requested to heed accuracy. When asked to list three prominent departments and their focus areas, the model gave sensible answers aligned with our knowledge. When asked to identify 10 articles that best represent the T4 stage, the model returned appropriate answers that demonstrated an understanding of the implied emphasis on population health and widespread implementation. Lastly, the model provided well-organized advice with annotated experts and strategies, when asked to recommend collaborators for doing AI research in women's health. DISCUSSION/SIGNIFICANCE OF IMPACT: Based on a literature dataset, the general-purpose AI model delivered impressive results when asked to fulfill tasks that involved information extraction, summarization, reference to external knowledge, and strategy consultation. The ongoing work aims to enhance scalability and accessibility of the tool while adding more rigorous evaluations.

56

Automated assessment of selective motor control in preterm infants using computer vision*

Divya Joshi[1], Colleen Peyton[2] and James Cotton[1]
[1]Northwestern University/Shirley Ryan Ability Lab and [2]Northwestern University

OBJECTIVES/GOALS: Cerebral palsy (CP) results in impaired selective motor control (SMC), or the ability to move joints

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 18 Nov 2026 at 09:46:20, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/cts.2026.10285

independently. This study leverages computer vision algorithms to create a scalable, low-cost, and accessible tool that automatically quantifies SMC behavior in infants born preterm, allowing for earlier screening and intervention for CP. METHODS/STUDY POPULATION: In single-camera videos of 600 15-week-old infants, all of whom have standardized clinical SMC scores available, human pose estimation algorithms will be used to derive 3D joint trajectories and kinematic metrics. These features will be used to train and cross-validate a machine learning model aimed at automating clinical scoring of SMC from single-camera video recordings of infants. Then, this model will be used to automatically score the SMC in longitudinal single-camera videos of 47 preterm infants (collected every two weeks from birth until 16 weeks of age). The SMC score trajectories between infants who eventually received a CP diagnosis and those who did not will be compared to establish an inflection point at which SMC development begins to diverge, establishing an optimal timepoint for intervention. RESULTS/ANTICIPATED RESULTS: Preliminary results show that 3D angles of 22 joints across the infant's body can be estimated in each frame of a 1 to 3 minute single-camera video. We anticipate that the error between the 3D angle estimated from a custom fitted kinematic model and the 2D angle calculated using human pose estimation measures will be less than 5 degrees for all joints measured. With a dataset of over 1,500 videos from over 600 infants, we expect that the trained machine learning model will achieve a sensitivity and specificity of over 90% in predicting the clinical SMC score. Finally, we expect that the automated, longitudinal SMC score trajectories will begin to significantly deviate in infants with a later CP diagnosis, from those without, at around term age (40 weeks gestational age), based on recent clinical findings. DISCUSSION/SIGNIFICANCE OF IMPACT: We provide the first automated, quantitative analysis of SMC from video recordings that can be used within real-world settings such as the NICU, at home, or at daycare. This work will result in a cost-effective, precise, smartphone-accessible tool that empowers earlier, more equitable diagnosis and treatment planning.

57

Translational evaluation of multimodal artificial intelligence for dermatology triage

Nuran Golbasi[1], Olivier Elemento[2], Veronica Rotemberg[3] and Shari R. Lipner[4]

[1]Weill Cornell Medical College, New York; [2]Englander Institute for Precision Medicine, Institute for Computational Biomedicine, Weill Cornell Medicine, New York; [3]Dermatology Service, Memorial Sloan Kettering Cancer Center, New York and [4]Department of Dermatology, Weill Cornell Medicine, New York

OBJECTIVES/GOALS: To evaluate the translational reliability, reproducibility, diagnostic performance, and subgroup equity of multimodal artificial intelligence (AI) models for dermatology triage across multiple model platforms. METHODS/STUDY POPULATION: Limited access to dermatology expertise delays diagnosis and care, motivating development of multimodal AI systems that integrate clinical images with patient data for triage. We assembled 200 biopsy-confirmed PAD-UFES-20 lesions (melanoma, keratinocyte carcinoma, benign) with paired images and metadata, prioritizing demographic balance. Six multimodal AI models (GPT-5, GPT-5-mini; Gemini 2.5 Pro, Gemini 2.5 Flash; Claude Sonnet-4, Claude Opus-4) analyzed these lesions with identical prompts predicting diagnostic probabilities, triage (urgent vs routine), and rationale. Outcomes included sensitivity, specificity, AUROC, F1, and subgroup equity. Model rationales were reviewed for interpretability, and subset re-prompting tested reproducibility for translational robustness. RESULTS/ANTICIPATED RESULTS: Across six models, sensitivity range was 0.89–1.00, specificity 0.21–0.65, AUROC 0.77–0.87, and F1 scores 0.72–0.81. GPT-5 achieved the most balanced performance (0.92 sensitivity, 0.65 specificity, AUROC 0.87, F1 0.81), while Gemini 2.5 Pro and Flash reached perfect sensitivity but low specificity (0.21–0.25). Claude Sonnet-4 showed near-perfect sensitivity (0.99) but over-called benign cases (0.24 specificity), while Opus-4 had the lowest sensitivity (0.89). Urgent triage aligned with dermatologist biopsy patterns (87–97%), and sensitivity was consistent across sex and skin type ($p \geq 0.29$). Subset re-prompting produced similar results, supporting reproducibility. Model rationales reflected dermatologic reasoning, supporting interpretability, and translational readiness. DISCUSSION/SIGNIFICANCE OF IMPACT: Multimodal AI models showed balanced diagnostic performance for dermatology triage, with platform-specific trade-offs between sensitivity and specificity. Subgroup equity, interpretable rationales, and subset reproducibility define key elements for reliable translation into dermatology workflows and prospective validation.

60

Post-deployment artificial intelligence model monitoring, evaluation, and intervention in health systems: A scoping review for guidelines for AI model report and the literature

Chenyu Gai[1], Shauna M. Overgaard[2], Joshua W. Ohde[2], Momin M. Malik[2], Hanyin Wang[2], Matthew J. Spiten[2], Deepak K. Sharma[2], Chung Il Wi[2] and Young Juhn[2]

[1]Mayo Clinic and [2]Mayo Clinic, Rochester, Minnesota, USA

OBJECTIVES/GOALS: This scoping review aims to synthesize current literature on post-deployment monitoring of AI-enabled digital health solutions within clinical practice. Findings identify existing approaches and gaps that inform guidance for post-deployment monitoring in clinical practice. METHODS/STUDY POPULATION: We conducted a scoping review in accordance with PRISMA-ScR guidelines to characterize the current landscape of post-deployment monitoring in healthcare systems. A PubMed search targeted peer-reviewed articles in English published between 2015 and mid-2025, including text or MeSH terms on 1) health system/hospital; 2) artificial intelligence; 3) post-deployment; and 4) evaluation/monitoring. We performed a thematic analysis to identify common challenges, gaps, and opportunities in AI oversight. Additionally, we reviewed guidelines addressing post-deployment AI monitoring. All analyses were conducted using Rayyan.ai and Microsoft Word. RESULTS/ANTICIPATED RESULTS: Among the six studies included after the full-text review, five provide recommendations to ensure transparency, safety, and model performance.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 18 Nov 2026 at 09:46:20, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/cts.2026.10285