

Artificial Intelligence in Clinical Medicine: Generative and Interactive Systems at the Human-Machine Interface

Sajjad Fouladvand

*Real-World Evidence and Advanced Analytics, Johnson and Johnson
Brisbane, CA, USA
Email: Sfouladv@its.jnj.com*

Emma Pierson

*Jacobs Technion-Cornell Institute, Cornell Tech
New York City, NY, USA
Email: ep432@cornell.edu*

Ivana Jankovic

*Oregon Health and Sciences University
Portland, OR, USA
Email: jankovii@ohsu.edu*

David Ouyang

*Department of Cardiology, Cedars-Sinai Medical Center
Los Angeles, CA, USA
Email: david.ouyang@cshs.org*

Jonathan H. Chen

*Department of Medicine and Center for Biomedical Informatics Research, Stanford University
Stanford, CA, USA
Email: jonc101@stanford.edu*

Roxana Daneshjou

*Biomedical Data Science, Stanford University
Stanford, CA, USA
Email: roxanad@stanford.edu*

Artificial Intelligence (AI) models are substantially enhancing the capability to analyze complex and multi-dimensional datasets. Generative AI and deep learning models have demonstrated significant advancements in extracting knowledge from unstructured text, imaging as well as structured and tabular data. This recent breakthrough in AI has inspired research in medicine, leading to the development of numerous tools for creating clinical decision support systems, monitoring tools, image interpretation, and triaging capabilities. Nevertheless, comprehensive research is imperative to evaluate the potential impact and implications of AI systems in healthcare.

At the 2024 Pacific Symposium on Biocomputing (PSB) session entitled “Artificial Intelligence in Clinical Medicine: Generative and Interactive Systems at the Human-Machine Interface”, we spotlight research that develops and applies AI algorithms to solve real-world problems in healthcare.

Keywords: Artificial Intelligence, clinical medicine, decision support systems.

1. Introduction

Recent progress in AI has led to the development of advanced large language models (LLMs), image, genomic and tabular data analysis tools (Huang et al., 2023; Movva et al., 2023, 2023; Omiye et al., 2023; OpenAI, 2023; Singhal et al., 2023; Tate et al., 2023; Wehbe et al., 2023). Leveraging these AI models for real-world biomedical data analysis is critical for enhancing diagnostic accuracy, predicting patient outcomes, and personalizing treatment plans, ultimately contributing to improved patient care and health outcomes. However, systematic evaluation of the potentials and limitations of AI algorithms within the medical domain is crucial to ensure the efficacy, safety, and reliability of AI-driven healthcare solutions and interventions (Wornow et al., 2023).

Here, we highlight the accepted submissions for the Artificial Intelligence in Clinical Medicine: Generative and Interactive Systems at the Human-Machine Interface session at the Pacific Symposium on Biocomputing (PSB) 2024. A goal of this session is to showcase research that has identified a clinical need that can be addressed by AI methods. Accepted submissions include use cases of using generative and classical AI models for analyzing different clinical data modalities and for a variety of applications such as answering medical questions, medical image analysis, clinical note analysis, cognitive monitoring, digital twins, and other decision support systems.

2. Artificial Intelligence in Clinical Medicine

2.1. Medical text and clinical notes analysis

There have recently been numerous successful applications of LLMs in ingesting medical text and clinical notes to extract vital information and insights for enhanced patient care. Lozano et al. (2024) proposed Clinfo.ai: an open-source retrieval-augmented LLM system for answering medical questions using scientific literature. The authors evaluated Clinfo’s performance (along with the performance of other question-answering systems, which the proposed method improves on) on a benchmark the authors made publicly available. Systems like this highlight the potential of large language models to help clinicians stay abreast of the enormous (and growing) medical literature. Jiang et al. (2024) proposed VetLLM, a large language model for predicting diagnosis from veterinary notes. They evaluated whether LLMs can be used to extract diagnoses from

unstructured veterinary notes. This approach can more easily facilitate broad veterinary research given that previous work often relies on customized, specialized models for each diagnosis. The paper revealed that, even without fine-tuning, open-source LLMs like Alpaca-7B show promising performance in diagnosis extraction tasks; performance is further improved when the model is fine-tuned on datasets of veterinary notes.

Pakhomov et al. (2024) proposed a conversational agent for early detection of neurotoxic effects of medications through automated intensive observation. This paper presents an AI system for monitoring cognitive symptoms of neurotoxicity which can occur in response to some immunotherapies. The system, a conversational agent, conducts a cognitive assessment over the phone including both spontaneous speech and neurocognitive tests. The authors present the results of a pilot study. Such systems have the potential to allow for intensive monitoring of patients while reducing the burden on them and medical staff (since automated monitoring can be conducted while the patient remains at home).

2.2. Medical image analysis

Another compelling avenue where AI has shown promising results is in the realm of medical image processing. This domain has witnessed a remarkable transformation, with AI algorithms now capable of efficiently analyzing a wide range of medical images, including ultrasound, X-rays, MRI scans, and CT scans, to yield faster and more accurate diagnoses. Duffy et al. (2024) used convolutional neural networks (CNNs) to evaluate the performance of AI models on 2D and 3D cardiac ultrasound datasets. Generally recorded as 2D video data, newer ultrasound transducers allow the collection of 3D data that can be post processed into standard 2D view videos. Using previously published CNNs for echocardiography (Ouyang et al., 2020), Duffy et al. showed that biases in 2D data (foreshortening and off axis views) can be simulated from the 3D data and have important impacts on model output.

Li et al. (2024) proposed BrainSTEAM, a practical pipeline for connectome-based fMRI analysis towards subject classification. This work addressed the overfitting problem in Graph Neural Networks (GNNs) used for analyzing structured network data. BrainSTEAM uses a spatio-temporal module that includes an EdgeConv GNN model, an autoencoder, and a strategy to dynamically segment time series signals, construct correlation networks, capture regions of interest (ROIs) connectivity structures, denoise data, and enhance model training. BrainSTEAM was evaluated on two real-world neuroimaging datasets, ABIDE for autism prediction and HCP for gender prediction, showing superior performance compared to existing models. This framework is potentially applicable to other studies for connectome-based fMRI analysis, promising enhanced reliability for clinical applications. Finally, recognizing the variation in human-quantified

phenotypes, Vukadinovic et al. (2024) show that different ways of assessing left ventricular ejection fraction, including variation within the range of clinician-to-clinician variability, can cause significant impact on downstream analyses, including genome wide association studies, where less precise measurements have a substantial impact on signal for genetic loci. Compared with sample size variation, 1% less precision in measurements resulted in the equivalent loss of power as a 10% decrease in cohort sample size.

2.3. Neurobiology and cognitive function

Prantzos et al. (2024) presented MaTiLDA, which serves as an integrated machine learning and topological data analysis platform for brain network dynamics. Brain activity is recorded via electroencephalograms (EEGs); however, analyzing large volumes of recordings can be difficult. They introduced and publicly shared MaTiLDA to enable the use of machine learning with topological data analysis on EEG data. They then showed how their platform could be used to analyze EEG data from neurological disorders such as epilepsy.

Yang et al. (2024) showed that DNNs on brain MRI images can be used to detect and distinguish between normal subjects and subjects with cognitive impairments like Alzheimer's disease. Javedani Sadaei et al. (2024) proposed Zoish: a novel feature selection approach leveraging Shapley additive values for machine learning applications in healthcare. They present a feature selection python package leveraging Shapley additive values to simplify feature selection for a variety of healthcare prediction tasks. As an illustrative example, Zoish was applied to a predictive model on Parkinson's progression as measured by the Montreal Cognitive Assessment (MOCA) and showed not only greater predictive performance overall but also improved interpretability compared to another feature selection method. As AI models attempt to move away from the "black box", tools such as Zoish can help clinicians better understand how the models produce predictions

2.4. Human-machine interface

Moore et. al. (2024) proposed SynTwin: a graph-based approach for predicting clinical outcomes using digital twins derived from synthetic patients. SynTwin introduces a novel methodology for generating and utilizing digital twins for clinical outcome prediction in precision medicine. The approach begins by estimating the distance between subjects based on their features, and then uses these distances to construct a network. Communities of subjects are defined, and a population of synthetic patients is generated. Digital twins, selected from this synthetic patient population, are used to enhance the prediction of clinical endpoints. When applied to a population-based cancer registry, the SynTwin approach significantly improved the prediction of mortality compared to

using real data alone, demonstrating the potential of this method in advancing precision medicine efforts. Patel et. al. (2024) proposed optimizing computer-aided diagnosis with cost-aware deep learning models. They propose a deep learning computer-aided diagnosis system to address the common situation in healthcare in which a false negative is more serious than a false positive. Whereas traditional computer-aided diagnosis systems penalize both types of misclassification equally, the cost-aware neural net model described here shows how using cost as a hyperparameter can boost sensitivity while largely maintaining overall accuracy.

3. Conclusion

Submissions accepted at the Artificial Intelligence in Clinical Medicine: Generative and Interactive Systems at the Human-Machine Interface session underscore the expanding role of AI in clinical medicine. The array of studies, spanning from advancements in AI-driven medical text and clinical notes analysis to breakthroughs in medical image processing, neurobiology, and human-machine interfaces, highlights the potential of generative and classical AI to improve healthcare. The consistent theme across all submissions is the emphasis on practical, real-world applications, showing AI's capability to enhance diagnostic accuracy, monitor cognitive symptoms, analyze diverse data types, and augment clinical decision-making processes. Despite these advancements, the need for identifying clinical problems and ongoing evaluation and assessment of AI technologies in healthcare to ensure their safety, efficacy, and reliability remains paramount. The works presented herein contribute significantly to this ongoing dialogue, showcasing both the possibilities and the remaining challenges in integrating AI into the healthcare landscape.

References

- Duffy, G., Christensen, K., & Ouyang, D. (2024). Leveraging 3D Echocardiograms to Evaluate AI Model Performance in Predicting Cardiac Function on Out-of-Distribution Data. *Pacific Symposium on Biocomputing (PSB)*.
- Huang, Z., Bianchi, F., Yuksekgonul, M., Montine, T. J., & Zou, J. (2023). A visual–language foundation model for pathology image analysis using medical Twitter. *Nature Medicine*, 29(9), Article 9. <https://doi.org/10.1038/s41591-023-02504-3>
- Javedani Sadaei, H., Loguercio, S., & Shafiei Neyestanak, M. (2024). Zoish: A Novel Feature Selection Approach Leveraging Shapley Additive Values for Machine Learning Applications in Healthcare. *Pacific Symposium on Biocomputing (PSB)*.
- Jiang, Y., Irvin, J. A., Ng, A. N., & Zou, J. (2024). VetLLM: Large Language Model for Predicting Diagnosis from Veterinary Notes. *Pacific Symposium on Biocomputing (PSB)*.
- Li, A., Yang, Y., Cui, H., & Yang, C. (2024). BrainSTEAM: A Practical Pipeline for Connectome-based fMRI Analysis towards Subject Classification. *Pacific Symposium on Biocomputing (PSB)*.

- Lozano, A., Fleming, S. L., Chiang, C.-C., & Shah, N. (2024). Clinfo.ai: An Open-Source Retrieval-Augmented Large Language Model System for Answering Medical Questions using Scientific Literature. *Pacific Symposium on Biocomputing (PSB)*.
- Moore, J. H., Li, X., Chang, J. H., Tatonetti, N. P., Theodorescu, D., Chen, Y., Asselbergs, F. W., Venkatesan, M., & Wang, Z. P. (2024). SynTwin: A graph-based approach for predicting clinical outcomes using digital twins derived from synthetic patients. *Pacific Symposium on Biocomputing (PSB)*.
- Movva, R., Balachandar, S., Peng, K., Agostini, G., Garg, N., & Pierson, E. (2023). *Large language models shape and are shaped by society: A survey of arXiv publication patterns* (arXiv:2307.10700). arXiv. <https://doi.org/10.48550/arXiv.2307.10700>
- Omiye, J. A., Gui, H., Rezaei, S. J., Zou, J., & Daneshjou, R. (2023). *Large language models in medicine: The potentials and pitfalls* (arXiv:2309.00087). arXiv. <https://doi.org/10.48550/arXiv.2309.00087>
- OpenAI. (2023). *GPT-4 Technical Report* (arXiv:2303.08774). arXiv. <https://doi.org/10.48550/arXiv.2303.08774>
- Ouyang, D., He, B., Ghorbani, A., Yuan, N., Ebinger, J., Langlotz, C. P., Heidenreich, P. A., Harrington, R. A., Liang, D. H., Ashley, E. A., & Zou, J. Y. (2020). Video-based AI for beat-to-beat assessment of cardiac function. *Nature*, *580*(7802), Article 7802. <https://doi.org/10.1038/s41586-020-2145-8>
- Pakhomov, S., Solinsky, J., Michalowski, M., & Bachanova, V. (2024). A Conversational Agent for Early Detection of Neurotoxic Effects of Medications through Automated Intensive Observation. *Pacific Symposium on Biocomputing (PSB)*.
- Patel, C., Wang, Y., Ramaraj, T., Tchoua, R., Furst, J., Raicu, D. (2024). Optimizing Computer-Aided Diagnosis with Cost-Aware Deep Learning Models. *Pacific Symposium on Biocomputing (PSB)*.
- Prantzalos, K., Upadhyaya, D., Shafiabadi, N., Fernandez-BacaVaca, G., Gurski, N., Yoshimoto, K., Sivagnanam, S., Majumdar, A., Sahoo, S. (2024). MaTiLDA: An Integrated Machine Learning and Topological Data Analysis Platform for Brain Network Dynamic. *Pacific Symposium on Biocomputing (PSB)*.
- Singhal, K., Azizi, S., Tu, T., Mahdavi, S. S., Wei, J., Chung, H. W., Scales, N., Tanwani, A., Cole-Lewis, H., Pfohl, S., Payne, P., Seneviratne, M., Gamble, P., Kelly, C., Babiker, A., Schärli, N., Chowdhery, A., Mansfield, P., Demner-Fushman, D., ... Natarajan, V. (2023). Large language models encode clinical knowledge. *Nature*, *620*(7972), Article 7972. <https://doi.org/10.1038/s41586-023-06291-2>
- Tate, S., Fouladvand, S., Chen, J. H., & Chen, C.-Y. A. (2023). The ChatGPT therapist will see you now: Navigating generative artificial intelligence's potential in addiction medicine research and patient care. *Addiction*. <https://doi.org/10.1111/add.16341>
- Vukadinovic, M., Renjith, G., Yuan, V., Kwan, A., Cheng, S. C., Li, D., Clarke, S. L., & Ouyang, D. (2024). Impact of Measurement Noise on Genetic Association Studies of Cardiac Function. *Pacific Symposium on Biocomputing (PSB)*.

Wehbe, R. M., Katsaggleos, A. K., Hammond, K. J., Hong, H., Ahmad, F. S., Ouyang, D., Shah, S. J., McCarthy, P. M., & Thomas, J. D. (2023). Deep Learning for Cardiovascular Imaging: A Review. *JAMA Cardiology*. <https://doi.org/10.1001/jamacardio.2023.3142>

Wornow, M., Xu, Y., Thapa, R., Patel, B., Steinberg, E., Fleming, S., Pfeffer, M. A., Fries, J., & Shah, N. H. (2023). The shaky foundations of large language models and foundation models for electronic health records. *Npj Digital Medicine*, 6(1), Article 1. <https://doi.org/10.1038/s41746-023-00879-8>

Yang, Y., Sathe, A., Schilling, K., Shashikumar, N., Moore, E., Dumitrescu, L., Pechman, K. R., Landman, B. A., Gifford, K. A., Hohman, T. J., Jefferson, A. L., & Archer, D. B. (2024). A deep neural network estimation of brain age is sensitive to cognitive impairment and decline. *Pacific Symposium on Biocomputing (PSB)*.