Knowledge-Infused Learning for Developing a Mental Health Diagnostic Copilot in Healthcare Systems

Gerald Ketu Ndawula
University of Maryland, Baltimore County
geraldn1@umbc.edu

Abstract

Healthcare diagnostics, especially in underserved communities, faces critical gaps in accessibility and accuracy. African Americans experience significant disparities in mental health care, often receiving delayed or inadequate treatment. This research proposes a diagnostic copilot, an AI-powered assistant designed to work alongside healthcare professionals. Using Knowledge-Infused Learning (KIL) and multi-turn conversations, the system integrates clinical knowledge and patient input to deliver actionable, explainable diagnoses in real-time. By engaging with both patients and clinicians, the copilot aims to reduce disparities, enhance trust, and improve diagnostic accuracy in mental health care.

Introduction

The need for accessible, accurate, and real-time healthcare diagnostics is more urgent than ever, particularly in underserved communities. African Americans face significant disparities in mental health disorder diagnoses and treatment. According to the CDC, African Americans are 20% more likely to suffer from serious mental health issues than the general population, but less than one-third receive the care they need (CDC 2021). The COVID-19 pandemic further highlighted gaps in healthcare accessibility, leaving millions without the care they require (CDC 2021).

My research proposal is to develop a diagnostic copilot, an AI-powered assistant designed to work alongside healthcare professionals, particularly in mental health. Unlike current AI systems that require clinicians to refine recommendations offline, the diagnostic copilot will seamlessly integrate clinical knowledge and patient input in an end-to-end pipeline. This ensures that the AI model is grounded in medical knowledge, is instruct-able by clinicians, and provides actionable and explainable outcomes. With the growing capabilities of Generative AI (Gen AI) in multi-turn conversations and patient monitoring, now is an opportune time to integrate such systems into healthcare. This approach aims to widen healthcare participation by enhancing accessibility for underserved communities, ensuring they receive timely, accurate care.

Background

AI systems in healthcare have traditionally focused on offering suggestive outputs based on large datasets of patient records. Solutions like IBM's Watson for Oncology, Google's DeepMind Health, and Google's Med-PaLM provide treatment recommendations, but they are often static, failing to adapt to the evolving nature of patient conditions. This limits their effectiveness, especially for conditions like mental health disorders, where continual reassessment based on patient feedback is critical (Topol 2019).

Additionally, current AI diagnostic systems heavily rely on large datasets, which often overlook underrepresented populations, leading to inaccurate diagnoses for marginalized communities (Gianfrancesco et al. 2018). African Americans, for instance, are disproportionately affected by mental health diseases, but existing models are not well-suited to address the nuances of their healthcare needs. These models fail to engage in real-time, multi-turn conversations with patients, limiting their adaptability to changing symptoms.

Recent research on Knowledge-Infused Learning (KIL) demonstrates that embedding domain-specific knowledge, such as medical ontologies, directly into AI models improves interpretability and adaptability, particularly in low-data scenarios (Chandak, Barzilai, and Cheng 2021). However, these systems still lack real-time interaction with patients, which is crucial for accurate diagnosis. Further, most AI systems focus on suggestive outputs without integrating clinical knowledge end-to-end into the decision-making process.

Approach

The proposed diagnostic copilot will address the limitations of current AI healthcare models by providing a dynamic, multi-turn conversational agent that works alongside healthcare professionals. This AI system will engage patients in real-time, interactive discussions to clarify symptoms, gather additional insights, and refine diagnostic suggestions. Such an approach is particularly valuable for mental health disorder management, where ongoing patient input is essential for accurate diagnosis. The copilot will ensure seamless clinician involvement, grounding all decisions in clinical knowledge, thus maintaining trustworthiness and explainability throughout the process.

Knowledge Extraction and Integration

The diagnostic copilot will extract relevant medical knowledge from trusted sources such as the Systematized Nomenclature of Medicine – Clinical Terms (SNOMED CT) and the Unified Medical Language System (UMLS). These ontologies provide structured and hierarchical information on symptoms, diagnoses, and treatments. A comprehensive knowledge graph will be built to link symptoms to their potential diagnoses (Jiang, Liu, and Gao 2020).

Once the relevant knowledge has been extracted, it will be embedded into decision trees and ensemble models. This integration will be achieved by modifying the splitting criteria of the models to align with medical hierarchies, such as pathways that link symptoms to specific diseases. Incorporating domain expertise ensures that the learning process follows a structured clinical reasoning model, grounded in established clinical practices (Jiang, Liu, and Gao 2020).

Model Training

The diagnostic copilot will be trained using transfer learning on datasets such as ESCON (Liu, Zhang, and Liu 2020), IHMI (Yang and Huang 2024), SMHD (Cohan et al. 2018), and Primate (Gupta et al. 2022). These datasets provide a diverse set of electronic health records and conversations related to mental health that reflect real-world patient cases. After the initial training phase, the models will be fine-tuned using knowledge-enriched datasets, derived from the knowledge graph developed in the extraction phase. This ensures that the copilot remains generalizable across different patient populations and healthcare environments and remains grounded on clinical concepts.

Implementing Multi-turn Feedback

Multi-turn feedback will be a core feature of the diagnostic copilot, allowing it to refine its diagnostic output continuously based on real-time patient input. By leveraging natural language processing (NLP) techniques, the copilot will engage patients in interactive, conversational loops. This system will ask follow-up questions, clarify reported symptoms, and gather additional details to create a more comprehensive diagnostic picture. For instance, if a patient reports new symptoms during the interaction, the copilot will integrate this information into its decision-making process, updating its predictions in real-time.

Evaluation Metrics

1. Clinical Entity Overlap Score: This metric evaluates the copilot's ability to recognize clinically relevant entities. The overlap between the concepts the model identifies and clinical concepts from the knowledge graph will be measured to ensure alignment with medical knowledge.

2. Logical Coherence and Semantic Relatedness: Using a pretrained RoBERTa model for Natural Language Inference (NLI), the system will be assessed for logical coherence in conversations. Additionally, cosine similarity and BERT score will measure semantic relatedness to ensure the conversational agent's responses remain contextually accurate (Chen, Chen, and Sun 2020).

3. Instructability (Clinician Feedback): The copilot's ability to adapt to clinician instructions will be evaluated using Direct Preference Optimization (DPO) or Kahneman-Tversky Optimization. This will assess how effectively the AI agent captures clinician input and improves diagnostic accuracy based on reinforcement learning with human feedback.

Broad Impact

The introduction of a diagnostic copilot into healthcare could revolutionize clinical AI systems, particularly in mental health disorders. By actively engaging with patients through multi-turn conversations and being instructable by clinicians, the copilot ensures that underserved communities, such as African Americans, receive timely and accurate care. This interactive system will allow for the continuous refinement of diagnoses based on real-time patient input, helping to mitigate disparities in care delivery.

By reducing training times and enhancing the adaptability of models to new clinical scenarios, the diagnostic copilot will make AI-driven healthcare solutions more scalable and reliable. The copilot's ability to provide clear, traceable reasoning for its diagnostic recommendations will foster trust between patients and clinicians, helping to bridge the gap in healthcare accessibility and equality.

Conclusion

This research seeks to develop a diagnostic copilot powered by Knowledge-Infused Learning, designed to enhance AI-driven healthcare diagnostics by improving accuracy, explainability, and trustworthiness. By leveraging clinical knowledge, engaging in real-time multi-turn interactions, and being instructable by clinicians, this copilot aims to provide more accurate diagnoses, reduce training times, and ensure reliable healthcare applications for underserved communities.

Acknowledgments

The author would like to thank Dr Manas Gaur, the Knowledge Infused AI and Inference Lab at University of Maryland Baltimore County and the University of Maryland, Baltimore County, for their support and the research mentors who provided invaluable guidance throughout this project.

References

CDC. 2021. Mental Health in Black Communities. Retrieved from https://www.cdc.gov/mentalhealth.

Chandak, S.; Barzilai, M.; and Cheng, J. 2021. Integrating medical ontologies into clinical decision support systems using knowledge-based AI models. *Journal of Healthcare Informatics Research*, 5(2): 177–190.

Chen, T.; Chen, D.; and Sun, L. 2020. Measuring Semantic Similarity in Health Conversations Using BERT. *IEEE Journal of Biomedical and Health Informatics*, 24(8): 2240–2250.

Cohan, A.; Desmet, B.; Yates, A.; Soldaini, L.; MacAvaney, S.; and Goharian, N. 2018. Self-reported Mental Health Diagnoses (SMHD) Dataset. In *Proceedings of the 27th International Conference on Computational Linguistics*, 1485–1497.

Gianfrancesco, M.; Tamang, S.; Yazdany, J.; and Schmajuk, G. 2018. Potential biases in machine learning algorithms using electronic health record data. *JAMA Internal Medicine*, 178(11): 1544–1547.

Gupta, S.; Agarwal, A.; Gaur, M.; Roy, K.; Narayanan, V.; Kumaraguru, P.; and Sheth, A. 2022. Process Knowledge for Depression Triage (Primate) Dataset. In *Proceedings of the 8th Workshop on Computational Linguistics and Clinical Psychology*, 112–123.

Jiang, H.; Liu, W.; and Gao, S. 2020. Enhancing Clinical Decision Support Systems with Knowledge-Based AI Models. *IEEE Access*, 8: 56041–56055.

Liu, Z.; Zhang, Q.; and Liu, Z. 2020. Emotional Support Conversation (ESCON) Dataset. *IEEE Transactions on Cognitive and Developmental Systems*, 12(2): 184–196.

Topol, E. 2019. High-performance medicine: The convergence of human and artificial intelligence. *Nature Medicine*, 25(1): 44–56.

Yang, K.; and Huang, M. 2024. Interpretable Mental Health Instruction (IHMI) Dataset. *IEEE Transactions on AI for Healthcare*.