Breakout Session 3: Track B

Shedding Light on the Black Box: Using Explainable AI to Enhance Clinical Research

Dr. Qing Zeng-Treitler (Moderator)
Director, Biomedical Informatics Center, George Washington University

Shedding Light on the Black Box: Using Explainable AI to Enhance Clinical Research

Use Explainable AI to Improve the Trust of and Detect the Bias of AI Models

Qing Zeng-Treitler PhD

Summary of the project and project goals

The parent grant was designed to test the hypotheses that higher cardiorespiratory fitness (CRF) is independently linked to a lower risk of incident AD/ADRD and to determine optimal levels of CRF assessed by a standardized exercise tolerance test (ETT), for minimizing AD/ADRD risk at both population and individual levels.

The supplement aims at advance the ethical development and use of AI/ML in biomedical and behavioral sciences using explainable AI methods.

Explainable AI (XAI)

- AI (with deep neural network models [DNN] as the prime examples) are emerging as an important method in biomedical research and delivery
- DNN models are black boxes to most clinicians and researchers
  - DNNs often have millions to billions of parameters
- Understanding and explaining their behaviors is the subject of ongoing research

HVAT

- We have designed a new Transformer-based DNN architecture, referred to as Hybrid Value-Aware Transformer (HVAT), which can jointly learn from longitudinal and non-longitudinal clinical data.

- On a case-control dataset, it achieved high performance in predicting
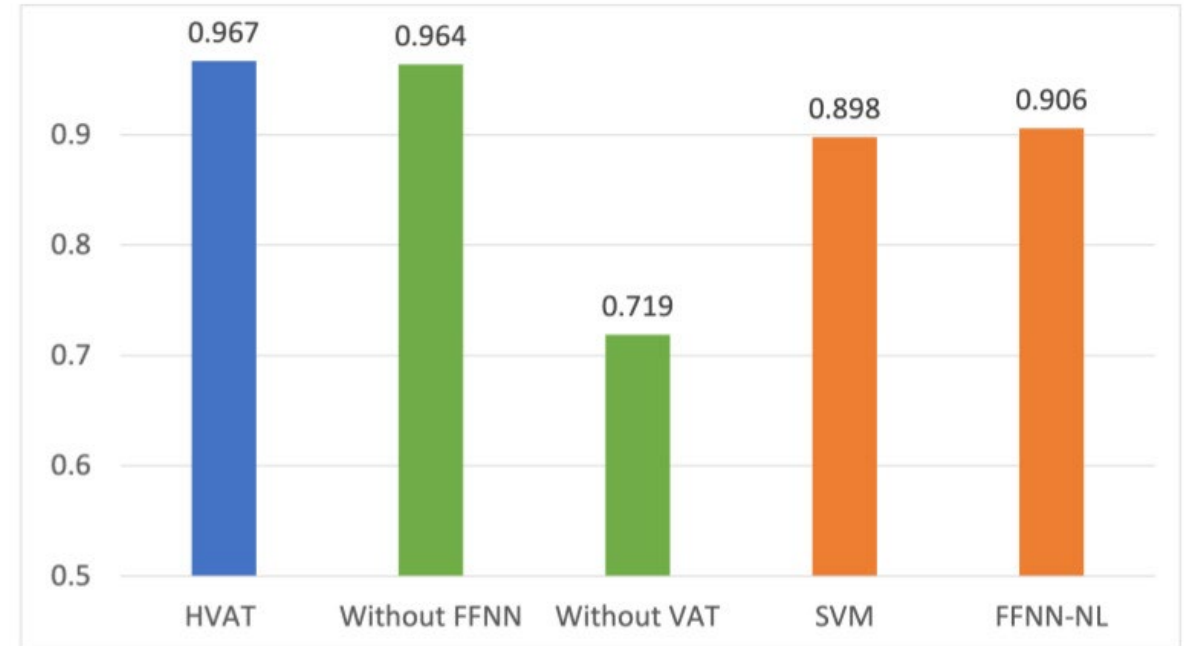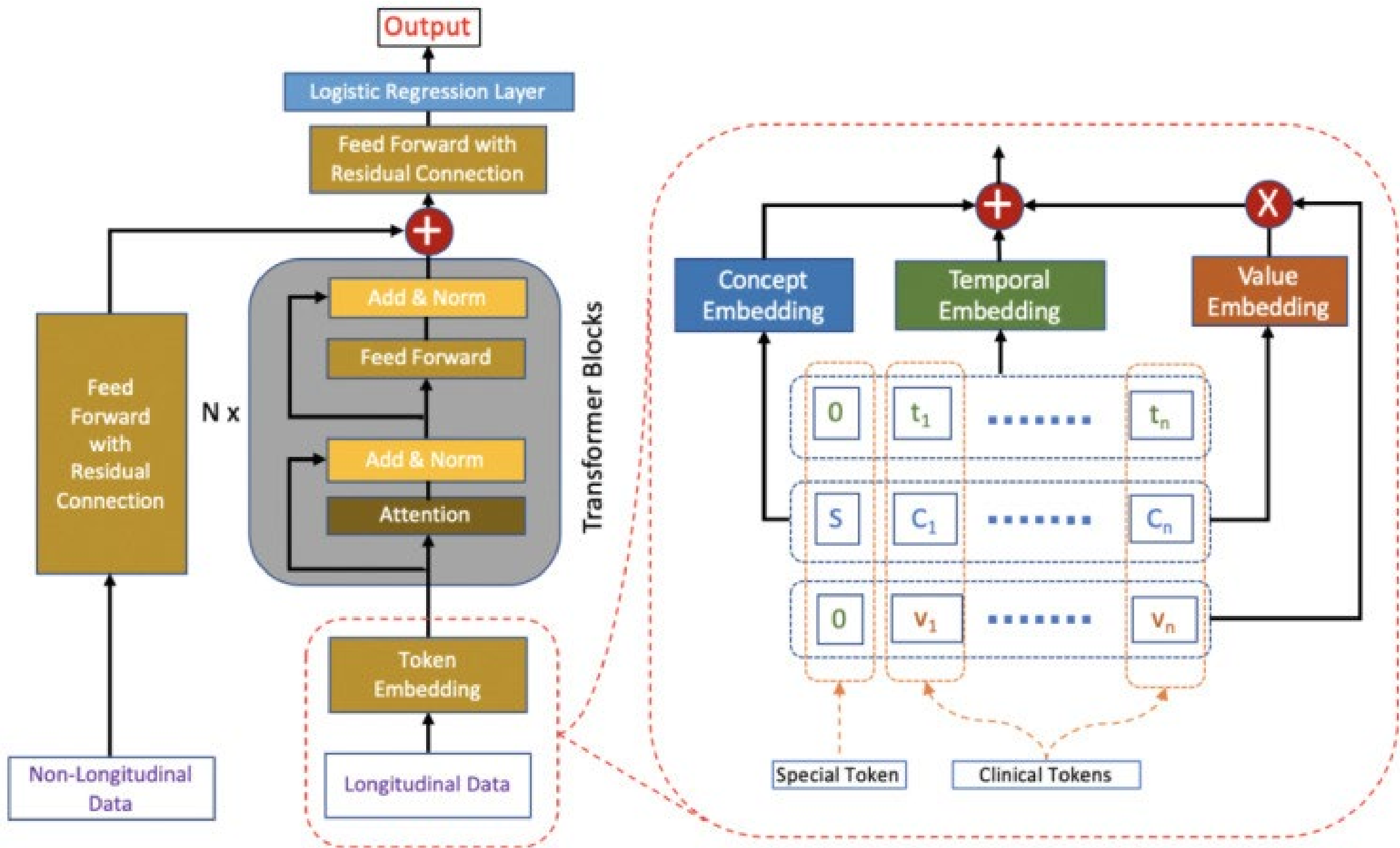
Assessing bias

Impact and interaction scores

- Impact scores quantify the impact of individual variables on the outcome. They generalize the coefficients of the predictor variables in linear logistic regression models to DNN models

- Interaction scores quantify the covariate interactions in DNN models predicting binary outcomes, which can be viewed as the generalizations of the interaction coefficients from regression models to DNN models.

Variable	Impact Score
AGE	2.198
TBI	1.342
DEPRESSION	1.160
STROKE	0.974
BMI	−0.738
CKD	0.693
RACE: Black (vs. White)	0.516

African Americans have a higher rate of increase in ADRD risk with age than white Americans. This is consistent with the discovery from a prior study by Guland et al. Another study by Chen et al. [26] discovered that increased BMI was associated with reduced risk of dementia. According to their results, obesity had a slightly stronger protective effect on African Americans than white cases. These discoveries are also in agreement with our results.

Variable 1	Variable 2	Interaction Score
AGE	RACE: Black (vs. White)	0.6891
BMI	RACE: Black (vs. White)	−0.5328
RACE: Black (vs. White)	ETHNICITY: Hispanic (vs. Non-Hispanic)	−0.2355
ATHRITIS	RACE: Black (vs. White)	−0.1875
AFIB	RACE: Black (vs. White)	0.1851
RACE: Black (vs. White)	AREA: Rural (vs. Urban)	−0.1512
CKD	RACE: Black (vs. White)	0.1474

Stakeholder Interviews

- We interviewed 17 individuals (3 clinicians, 3 patients, 3 bioethics experts, 2 IT experts, 3 healthcare administrators, and 3 policy makers)

- Main theses
  - XAI is very important topic to all stakeholders
  - They believe that transparency and comprehensibility are crucial for building trust and confidence in AI systems.
  - Many respondents emphasize the need for clear explanations of how AI arrives at its decisions.

Patient Perspective:
"But it actually would give me more confidence in the doctor to be able to diagnose it properly, just to be able to understand that they know what they're doing to be able to explain to me how AI is working to help me..."
"I think it is very important. It's the concept of transparency. AI is kind of this black box to most people, so not being able to understand everything, but understand a reasonable layman's term explanation of how things work..."

Clinician Perspective:
"I need a study that tells me that this AI system decision-making process is accurate. I need to understand that this thing is very accurate or else I'm going to use it and make the wrong diagnosis and wrong decisions."

"need to understand in layman's terms explanation of how things work and acknowledging that there can be biases and have to have good information going in to get an accurate answer coming out."

IT Expert Perspective:
"As a practitioner myself, there's a certain level of information that I would want to have, but I don't know if it's necessary for every clinician to have a detailed understanding of how these decisions are going to be made especially at the pace with which they're making decisions, it seems unrealistic"

Health Administrators Perspective:
"Form a subcommittee of various stakeholders like those you're interviewing here to be able to adjudicate the technologies that come through and make sure that they meet the standards that we think are appropriate for our culture and environment"
"There should be some human supervision to ensure that the results make sense. Many people will lose trust in the system if the explainability/results do not make sense"

Policymaker Perspective:

"The more that is explainable and objective, the more people are going to use and feel comfortable using it."

"If you are understanding how it works and how you can actually use it for our own advantages. We are able to streamline a lot of decision-making processes, whether it is developing the policy, implementing it."

"The potential risks are making it too "techy" and then you don't really get a buy-in the system or you don't let people to address the various issues. You will get a lot of feedback from patients, so do not make it complicated and make it simple. It needs to be understood from the patient and the doctor."

Ethics Expert Perspective:
[Explainable AI] "100% important. There needs to be an explanation of what databases the AI was taught about, what testing it's gone through, and what regulatory process it's been through. We need to know what's under the hood and how it reaches its decisions."

"More interpretability and testing. Transparency is knowing what the reasoning is. Accountability is a tricky one because it is difficult to know who to blame for errors in the AI. Just as if there were a faulty medical device where the manufacturer would have some liability. If you have a faulty AI algorithm that leads to a poor outcome for a patient, you'd expect there to be some liability for the developer of the algorithm"

"It's [transparency] very important because if you were able to rigorously test something to be confident in its accuracy and reliability and fairness, then you would feel more comfortable with deploying it. There would be much more confidence in the system if you know how it's coming to its outputs."

Challenges and how these were addressed

- Explaining an AI model to end users/stakeholders remains very challenging.

- We have been in dialogue with the bioethics expert on the project to look at explanation in the broader AI framework.

- For clinicians and healthcare consumers, explanation needs to be tied into transparency, bias, security, certainty/robustness to develop trust.

Related publications and presentations

- Shao Y, Cheng Y, Nelson SJ, Kokkinos P, Zamrini EY, Ahmed A, Zeng-Treitler Q. Hybrid Value-Aware Transformer Architecture for Joint Learning from Longitudinal and Non-Longitudinal Clinical Data. J Pers Med. 2023 Jun 29;13(7):1070. doi: 10.3390/jpm13071070. PMID: 37511683; PMCID: PMC10381142.

- Shao Y, Cheng Y, Nelson SJ, Kokkinos P, Zamrini EY, Ahmed A, Zeng-Treitler Q. Hybrid Value-Aware Transformer Architecture for Joint Learning from Longitudinal and Non-Longitudinal Clinical Data. medRxiv [Preprint]. 2023 Mar 14:2023.03.09.23287046. doi: 10.1101/2023.03.09.23287046. Update in: J Pers Med. 2023 Jun 29;13(7): PMID: 36993767; PMCID: PMC10055462.

- Mitigating AI Risk through Ethical Data Science. Workshop at AMIA 2024 Symposium

Future work

Visualization of impact and interaction

Explaining temporal relationship

Integrated approach to enhance trust