Stanford

Soumyadeep Roy

Postdoctoral Scholar, Biomedical Informatics

Curriculum Vitae available Online

Bio

BIO

I am a postdoctoral scholar at the Center for Biomedical Informatics Research of Stanford University, advised by Prof. Tina Hernandez-Boussard.

My primary area of research is natural language processing, with expertise in medical and healthcare applications. My research areas of interest are Foundation Models for Medicine, Generative AI, Text Summarization, and Efficient Pretraining.

I hold a PhD in Computer Science and Engineering from the Indian Institute of Technology Kharagpur, where I worked with Prof. Niloy Ganguly and Prof. Shamik Sural. Here, I was part of the Complex Networks Research Group (CNeRG). My PhD thesis is titled “Domain Adaptation for Medical Language Understanding”, where I developed novel domain adaptation techniques to effectively and efficiently adapt open-domain AI models to the medical domain.

In summary, I have six years of experience working with medical NLP data, which includes clinical trial registry data (2018-2021), medical forum questions (2020-2021), DNA sequence data (2021-2024), biomedical scientific literature (2023 - 2025), clinical data (2021-2023) and EHR clinical notes (2025). My medical AI research experience includes 2.5 years at L3S Research Germany collaborating with Hannover Medical School as well as a 7-month research internship at GE HealthCare Technology and Innovation Center (HTIC) in Bangalore, India. I also presented a tutorial on March 10, 2025 titled "Building Trustworthy AI Models for Medicine" at WSDM 2025 held in Germany.

In my free time, I like hiking, and playing chess or table tennis.

PROFESSIONAL EDUCATION

- Doctor of Philosophy, Indian Institute of Technology, Kharagpur (2025)
- Master of Science, Indian Institute of Technology, Kharagpur (2019)
- Bachelor of Technology, Maulana Abul Kalam Azad University of Technology (2017)

STANFORD ADVISORS

- Tina Hernandez-Boussard, Postdoctoral Faculty Sponsor

LINKS

- Personal Website: https://roysoumya.github.io/
- Google Scholar: https://scholar.google.com/citations?user=t3EF4i0AAAAJ&hl=en

Publications

PUBLICATIONS

- **Decision tree-based approach to robust Parkinson's disease subtyping using clinical data of the Michael J. Fox Foundation LRRK2 cross-sectional study.** *Frontiers in artificial intelligence*
  Roy, S., Krähe, S., Marschollek, M., Frieling, H., Ganguly, N., Wolff, D.
  2025; 8: 1668206

- **Building Trustworthy AI Models for Medicine: From Theory to Applications**
  Roy, S., Sundaram, S. S., Wolff, D., Ganguly, N., ACM
  ASSOC COMPUTING MACHINERY.2025: 1012-1015