NARRATIVE CV
SOUMYA BANERJEE

Personal Details

My research is in computational medicine, machine learning and data science applied to healthcare and modelling of biological systems. I build computer models to investigate problems related to human health.

I received my PhD in computer science from the University of New Mexico in USA. I am currently a senior research fellow and affiliated lecturer at the University of Cambridge. I develop explainable AI techniques and apply them to various domains like healthcare. I also teach courses in AI and supervise student projects. I use machine learning techniques and data science on electronic healthcare record data to help stratify patients and predict mortality. This has been published in *Nature Partner Journal Schizophrenia*.

I use machine learning techniques to solve problems in healthcare and biological systems. I work closely with clinicians and experimentalists to solve problems that can have immediate impact. These include the use of clinical data to prioritise patients for therapies [3,4] (collaboration with clinicians and experts in the pharmaceutical industry experts) and the building of mechanistic models of cell biology [5,6] (collaboration with experimentalists).

I also apply machine learning techniques in healthcare, especially in developing countries. I was invited to submit a correspondence on the potential effect of a shortage of the drug hydroxychloroquine in developing nations. This has been published in the journal *Lancet Rheumatology* [2].

One feature of my research is my inter-disciplinary collaborative work with clinicians, biologists, social scientists, and industry experts. I use machine learning to perform patient stratification in different diseases. One work has been published in *Nature Communications* [3] and another is published in *Gut* (*British Medical Journal* publishing group) [4]. I am also developing explainable machine learning techniques and applying them to mental health using data from electronic healthcare records [9] (published in *Nature Partner Journal Schizophrenia*).

A key feature of my work is interdisciplinarity and co-design with different stakeholders. In recent work, I designed AI algorithms for healthcare in close collaboration with patients [9,13,14]. This work was published in the *Patterns* journal (*Cell Press* publishing group) [13].

I have worked very closely with inter-disciplinary teams to design these solutions. I want to teach students how to collaborate with different stakeholders including clinicians, computer scientists, biologists, social scientists, and industry experts. My students will learn new interdisciplinary ways of working that straddle scientific disciplines.

I was successful in securing funding as part of an interdisciplinary application for the AI@CAM initiative at the University of Cambridge. This will develop new AI tools for mental health in collaboration with patients.

Contribution to the generation and flow of new ideas hypothesis tools and knowledge

I have contributed to a novel methodology of how to develop artificial intelligence tools in collaboration with non-technical stakeholders like patients who have lived experience or disease

[13,14]. I have also developed novel explainable artificial intelligence algorithms that can be explained to non-technical experts such as clinicians and patients [9].

I have contributed to interdisciplinary research in several different disciplines. These range from immunology [1], cell biology [5,6], gastroenterology [4] and psychiatry [9]. In all these different projects, I have worked very closely with computer scientists, clinicians, and patients. I have designed artificial intelligence algorithms in close collaboration with all these stakeholders. These projects have led to novel insights into different diseases.

I have spent the last 10 years of my career after my Ph.D. in acquiring these skills of interdisciplinary research and understanding how best to conduct research with experts from disciplines that are far removed from computer science. I have also trained students in interdisciplinary research and co-design research with patients.

I have also developed novel software which is publicly available (https://sites.google.com/site/neelsoumya/software). These software allow machine learning techniques to be used by experts in different disciplines. The techniques range from Bayesian machine learning techniques [1] to privacy preserving machine learning techniques [10,11,12]. I have also designed software that can be run on a web browser and can allow non-technical people (such as patients) to get an understanding of artificial intelligence.

I have co-organised patient and public involvement workshops that allow patients to understand artificial intelligence applied to healthcare. These workshops also allowed my students to understand how to communicate complex ideas to patients and how to involve them in research.

About from my methodological outputs in machine learning techniques applied to different disciplines, I have also developed approaches that suggest how in equities in health care can be exacerbated by global pandemics. During COVID, I published a paper in the journal *Lancet Rheumatology* that suggested how a surge in demand for hydroxychloroquine could compromise the health of people in malaria endemic regions in Asia and Africa who are dependent on these medications.

I have also developed teaching resources that explain difficult concepts in machine learning to students and lay audiences. I have made all these videos and my teaching material freely available for download on various repositories (https://sites.google.com/site/neelsoumya/teaching).

Contributions to research teams and development of others

Over my career, I have taught multiple courses on topics ranging from machine learning to complex systems, to immunology. I have taught in multiple universities and in summer schools. My contribution has been to demystify these complex techniques for students and early career researchers who are contemplating entering the field of artificial intelligence.

I am very passionate about making my education resources and tools available to the public even outside the universities I have worked in. To this end I have multiple lectures, videos, and course materials publicly available on YouTube and GitHub. These resources are publicly available for download and consumption by students in both developed and developing nations (https://sites.google.com/site/neelsoumya/teaching).

A big barrier to entry for students in the field of artificial intelligence is maths anxiety. To this end, I have worked with an educator and co-created educational resources that can explain complex topics in mathematics and artificial intelligence to students and lay audiences. These resources have been published and have received very favourable feedback from peer reviewed journals [15,16].

I have also worked in very large teams on interdisciplinary projects. I have mentored my junior colleagues and peers. I have been a role model for junior colleagues who were not sure about pursuing a career in academia.

I have also mentored my students and taught them how to participate in interdisciplinary research and involve patients in their research. I trained one of my students in a very challenging interdisciplinary research project with a psychiatrist. I also trained the student and involved him in a patient and public involvement workshop. In this workshop, the student explained his work to patients and elicited their response on our work. In this way, the student was trained in how to communicate his research in an interdisciplinary setting and how to communicate effectively with nontechnical stakeholders such as patients.

Contributions to the wider research and innovation community

My contributions to the wider research community include my role as a reviewer and being on the editorial board of a journal. I also regularly contribute educational resources to various journals that that are focused on best practises in education and outreach.

My contributions include various educational resources that aim to popularise mathematics and difficult topics in science. This is extremely important since mathematics anxiety is very common in many students. My educational resources aim to demystify the inner workings of a mathematician and make mathematics fun (recreational mathematics).

At my university, I have also organised workshops on academic writing. This is motivated by the fact that many early career researchers and students struggled with academic writing. This dissuades many talented researchers from taking up academia as a long-term career. I organised workshops that give practical tips on academic writing to students and early career researchers.

Exploitation of research

I have been part of medical research using artificial intelligence that has led to partnerships with prominent pharmaceutical companies. Based on our collaboration and published research, these pharmaceutical companies are now trying to develop new therapeutics in the field of the gut microbiome [3] and autoimmune diseases [4].

The participation of patients and the public in research in artificial intelligence has been a focus of my research programme. On projects in mental health, I have involved patients in artificial intelligence research. One of our outputs has also been a methodology of how to involve patients in artificial intelligence research in healthcare [13,14].

Personal statement

My long-term goals are to build a research group that investigates different facets of computational biology, machine learning, and clinical informatics applied to healthcare.

Over the last 20 years, I have worked in industry in various sectors (such as finance, healthcare, and supply chains) and in academia (in sectors such as immunology, cell biology and psychiatry). I enjoy collaborating with experts in different disciplines. I apply computational techniques to solve problems in different domains. I love being exposed to different problems in different disciplines that can be tackled using computers and artificial intelligence.

I have also trained many students in these skills under disciplinary collaboration and have always been embedded in different departments.

I am also very passionate about involving patients and the public in artificial intelligence research. I believe this can help in building AI systems that we can all trust and understand. I am also involved in creating educational and outreach resources to demystify complex scientific topics.

References

[1] Soumya Banerjee, Alan Perelson, Melanie Moses, Modelling the effects of phylogeny and body size on within-host pathogen replication and immune response, *Journal of the Royal Society Interface,* 14(136), 20170479, 2017

[2] Soumya Banerjee, Hydroxychloroquine: balancing the needs of LMICs during the COVID-19 pandemic, *Lancet Rheumatology*, 2(7):385-386, 2020

[3] Himel Mallick, Eric Franzosa, Lauren McIver, Soumya Banerjee, Alexandra Sirota-Madi, Alex Kostic, Clary Clish, Hera Vlamakis, Ramnik Xavier, Curtis Huttenhower. Predictive metabolomic profiling of microbial communities using amplicon or metagenomic sequences, *Nature Communications*, 10(1):3136, 2019

[4] Dominik Aschenbrenner, Maria Quaranta, Soumya Banerjee, et al. Deconvolution of monocyte responses in inflammatory bowel disease reveals an IL-1 cytokine network that regulates IL-23 in genetic and acquired IL-10 resistance, *Gut*, 2021

[5] Peng Liu, Abram Calderon, Georgios Konstantinidis, Jian Hou, Stephanie Voss, Xi Chen, Fu Li, Soumya Banerjee, Jan-Erik Hoffmann, Christiane Theiss, Leif Dehmelt, Yao Wen Wu, A bioorthogonal small-molecule switch system for controlling protein function in live cells, *Angewandte Chemie*, 53(38), 2014

[6] Melanie Graessl, Johannes Koch, Abram Calderon, Dominic Kamps, Soumya Banerjee, Tomáš Mazel, Nina Schulze, Jana Kathrin Jungkurth, Rutuja Patwardhan, Djamschid Solouk, Nico Hampe, Bernd Hoffmann, Leif Dehmelt, Perihan Nalbant. An excitable Rho GTPase signaling network generates dynamic subcellular contraction patterns, *Journal of Cell Biology*, 216(12), 2017

[7] Soumya Banerjee, J. Guedj, R. Ribeiro, M. Moses and A.S. Perelson, Estimating biologically relevant parameters under uncertainty for experimental within-host murine West Nile virus infection, *Journal of the Royal Society Interface*, 13(117), 20160130, 2016

[8] Soumya Banerjee, SJ Chapman. Influence of correlated antigen presentation on T cell negative selection in the thymus, *Journal of the Royal Society Interface*, 15(148):20180311, 2018

[9] Soumya Banerjee, Pietro Lio, Peter Jones, Rudolf Cardinal. A class-contrastive human-interpretable machine learning approach to predict mortality in severe mental illness, *NPJ Schizophrenia*, 7, 60, 2021

[10] Soumya Banerjee, Ghislain Sofack, Thosoris Papakonstantinou, Demetris Avraam, Paul Burton et al. dsSurvival: Privacy preserving survival models for federated individual patient meta-analysis in DataSHIELD, *BMC Research Notes*, 15(1):197, 2022

[11] Soumya Banerjee, Tom Bishop dsSurvival 2.0: Privacy enhancing survival curves for survival models in the federated DataSHIELD analysis system, *BMC Research Notes* 16, 98, 2023

[12] Soumya Banerjee, Tom Bishop dsSynthetic: Synthetic data generation for the DataSHIELD federated analysis system, *BMC Research Notes* 15(1):230, 2022

[13] Soumya Banerjee, Phil Alsop, Linda Jones, Rudolf Cardinal. Patient and public involvement to build trust in artificial intelligence: a framework, tools and case studies, *Patterns* 3(6):100506, 2022

[14] Soumya Banerjee, Sarah Griffiths. Involving patients in artificial intelligence research to build trustworthy systems, *AI & Society*, 2023

[15] Soumya Banerjee, Joyeeta Ghose, Tarakeswar Banerjee, Kalyani Banerjee Beauty of Life in Dynamical Systems: Philosophical Musings and Resources for Students, *Journal of Humanistic Mathematics* 13(2):427-444, 2023

[16] Soumya Banerjee, Ramanujan cab numbers: a recreational mathematics activity, *Journal of Humanistic Mathematics* 12(2):503-517, 2022