UNIVERSITY OF SURREY

AI, BIG DATA, AND PREGANANCY LOSS

WORKSHOP PROGRAMME

12 July 2023

OUR SPONSORS

IAS Institute of Advanced Studies

The Institute of Advanced Studies (IAS) at the University of Surrey sponsors workshops and Fellowships at the 'cutting edge' of science, engineering, social science and the humanities. Through this scheme the Institute fosters interdisciplinary collaborations and encourages a flow of international scholars to visit, enjoy their stay at Surrey and leave behind excellent ideas and innovations.

ias.surrey.ac.uk

People-Centred AI
UNIVERSITY OF SURREY

The Surrey Institute for People-Centred Artificial Intelligence is a pan-university institute bringing together Surrey's core AI-related expertise in vision, speech and signal processing, computer science, and mathematics, with its domain expertise across engineering and physical sciences, human and animal health, law and regulation, business, finance and the social sciences.

surrey.ac.uk/artificial-intelligence

Organising committee:
Professor Inga Prokopenko, Dr Adam Mahdi, Dr Ayse Demirkan (School of Biosciences and Medicine, University of Surrey) Dr Yevheniya Sharhorodska (IAS Fellow, Institute of Hereditary Pathology of National Academy of Medical Sciences of Ukraine)

Administrative support:
Ms Louise Jones (Institute of Advanced Studies) and Ms Inna Trush (Unitemps)

INTRODUCTION

The health of women's reproductive system is often overlooked, even though it affects many aspects of their lives. Women face more reproductive challenges as they get older and delay childbearing. Many women use in-vitro fertilisation to overcome these challenges, but this can also worsen some reproductive problems, such as pregnancy loss (PL). PL is when a pregnancy ends before the baby can survive outside the womb. It can happen early (miscarriage) or late (stillbirth) in pregnancy. PL can cause physical and mental harm to women, such as bleeding, infection, emotional distress, infertility and heart disease.

Many studies demonstrated contribution of variety of factors to susceptibility of PL but haven't benefited from the recent technological development and availability of large datasets to the same extent as other common characteristics. Identification of the predisposing genomic and other risk factors may help in determining the causes of PL and facilitate development of novel therapeutics and improved health interventions. Artificial intelligence (AI) and machine learning approaches could be implemented to predict PL. This workshop will explore the implementation of machine learning and other AI approaches to combine multiple individual health data characteristics, genomic, metabolomic, blood biochemistry and other data for prediction of women's reproductive health outcomes during pregnancy and development of prevention strategies for health systems. We will also explore the vision about the AI for PL target areas from parents who experienced PL.

Surrey institute for People-Centred AI has consolidated a strong AI expertise with a focus on healthcare research, including a strong track record in high-dimensional omics method development and application (Prof Prokopenko; Dr Kaakinen), metagenomics (Dr Demirkan), healthcare data fusion AI (Dr Mahdi), signal (Dr Kiskin) and image (Dr Dutta) processing. The group has access to three large-scale European biobanks (UK, Estonian, and FinnGen) with self-reported and electronic health record data on 622,000 women with information about pregnancies and their outcomes. This data will combined with the newly launched two (GPU- and CPU-based) HPC clusters enabling big data modelling. Moreover, our successful fellowship project of the Institute of Advanced Studies (IAS) in a reproductive health already helps us to acquire information and undertake research on PL outcomes to support the growing Surrey's expertise. During the IAS fellowship 2021-2022, we developed collaboration with companies (Deloitte, UK) and international institutions (Lviv, Ukraine and Tashkent, Uzbekistan) to expand our effort on the implementation on modern multidisciplinary approaches to PL prevention using AI. Integration of expertise and methods from different disciplines: epidemiological, statistical, data management, bioinformatics, human genomics and Multiomics of reproductive health using AI will support growing Surrey's leadership in multiomics AI for women's reproductive health and PL, specifically, to obtain external (UK and international) funding tackling prevention of PL.

PROGRAMME

WEDNESDAY 20 JULY

Innovation for Health building 02EFH01

(BST)

08.30 – 09.00 Registration

09.00 – 09.10 **Welcome** - Prof Inga Prokopenko, University of Surrey, UK

Session One
Chair: Prof Inga Prokopenko, University of Surrey, UK

09.10 – 10.00 **Pregnancy Loss and related outcomes – clinical perspectives from the UK** - Prof Ranjit Akolekar, Medway Fetal and Maternal Medicine Centre, Medway NHS Foundation Trust, UK

10.00 – 10.15 Refreshment Break

Session Two
Chair: Dr Ayse Demirkan, AI institute, University of Surrey, UK

10.15 – 11.15 **Prediction of Maternal-Fetal Health** - Dr Julia Zollner, Institute for Women's Health, University College London, UK

11.15 – 11.30 Bio-Break

Session Three: Big omics data research into perinatal maternal health
Chair: Dr Zhanna Balkhiyarova, University of Surrey, UK

11.30 – 12.00 **Microbiome and virome analyses for pregnancy loss prevention** - Dr Ayse Demirkan, University of Surrey, UK

12.00 – 12.30 **Ukrainian study of Recurrent pregnancy loss and UK biobank as genomic research resources** - Dr Yevheniya Sharhorodska, University of Surrey, UK & Institute of Hereditary Pathology, Lviv, Ukraine

12.30 – 13.30 Lunch and **Poster Viewing**

13.30 – 14.00 **One slide enlightening poster pitch session**
Chair: Dr Yevheniya Sharhorodska, University of Surrey, UK & Institute of Hereditary Pathology, Lviv, Ukraine

Session Four: Big data analytics: return of health results to people, AI for omics data, compute and TREs
Chair: Prof Inga Prokopenko, University of Surrey, UK

14.00 – 14.40 **The Estonian Biobank's MyGenome Portal: A comprehensive platform for return of results to over 200,000 biobank participants** - Dr Natalia Pervjakova, Estonian Biobank, Estonia

14.40 – 14.50 Bio-Break

14.50 – 15.50 **ChatGPT and Foundation models for health care and omics** - Dr Muhammad Awais, AI institute, FEPS, University of Surrey, UK

15.50 – 16.30 **Scaling precision medicine through federation; how the sum is greater than the parts** - Dr Joe Foster, Lifebit, UK

16.30 – 17.00 **Session Four: Discussion. Next steps for research using AI with next generation omics and screening for prevention of pregnancy loss**
Chair: Prof Inga Prokopenko & Dr Adam Mahdi, University of Surrey, UK

17.00 – 17.30 Refreshment Break and **Poster Viewing**

17:30 Workshop close

ABSTRACTS AND PARTICIPANTS

Professor Ranjit Akolekar

Professor Ranjit Akolekar is a Professor in Fetal Medicine and obstetrics at Medway NHS Foundation Trust, Kent. He has over 200 peer-reviewed research publications, more than 20,000 citations and an H-index of 68.

Research interests include prevention of stillbirths; with particular interest in the prediction of complications related to the placenta and umbilical cord and early prediction of pregnancy complications such as preeclampsia.

Pregnancy Loss and related outcomes – clinical perspectives from the UK

Prof Ranjit Akolekar, Medway Fetal and Maternal Medicine Centre, Medway NHS Foundation Trust, UK

The NHS England is one of the only organisations internationally that has a national policy for preventing stillbirths that is embedded across all maternity units in England and Wales. The Saving Babies Lives is a good example of the desire to prevent avoidable harm to support women and families to have good pregnancy outcomes. There is always more that can be done towards this ambition of preventing stillbirths and how latest evidence can be incorporated into routine clinical practice to prevent stillbirths and making this dreaded outcome a never event. This talk will discuss the latest translational research and evidence that can be incorporated into maternity services to provide women and families with care they deserve and that we strive to provide.

Dr Julia Zöllner

Dr Julia Zöllner is interested in the translational applications of OMICS technologies for the early detection and personalised management of women with gestational syndromes. She is interested in bridging the gap of ethnic disparity in genomic medicine.

Her current active research project is 'Interrogating Genes & Health', a long-term genomic medicine longitudinal population study of British Bangladeshi and British Pakistani adult (age 16+) volunteers and leading the bioinformatic data analysis, interpretation, and dissemination of findings in the following projects:

1. Developing a polygenic risk score (PRS) to predict gestational diabetes at booking in women of South Asian origin. Evaluating PRS in conjunction with clinical parameters.
2. To investigate the genetic cause of intrahepatic cholestasis of pregnancy in women of South Asian origin (rare & common variant analysis).
3. Curation of a maternal database at Genes and Health.

Prediction of Maternal-Fetal Health

Dr Julia Zollner, Institute for Women's Health, University College London, UK

The prediction of maternal-fetal health is of paramount importance for ensuring optimal pregnancy outcomes. Recent advancements in genetics and omics technology have revolutionised our understanding of the intricate interplay between maternal and fetal factors. The identification of genetic variants associated with adverse pregnancy outcomes has provided valuable insights into the underlying molecular mechanisms. Genome-wide association studies and whole-exome sequencing have identified several genetic markers associated with conditions such as preeclampsia, or gestational diabetes. The question remains whether Integration of these genetic markers into prediction models can improve risk assessment and personalised care for pregnant women. The talk will provide an overview of the role of genetics and the utilization of omics technology in predicting maternal-fetal health.

Dr Ayse Demirkan

After obtaining her PhD at Erasmus Medical Center, Rotterdam, the Netherlands, Dr. Demirkan pursued her research on statistical multi-omics, focusing on lipidomics, proteomics and transcriptomics of common conditions such as diabetes. Since 2018 her focus includes human microbiome and understanding its role in human traits and particularly in neurodegeneration. She is currently a senior -lecturer of section Statistical -multi-omics at University of Surrey and a core fellow of the People-centered AI institute.

Microbiome and virome analyses for pregnancy loss prevention

Dr Ayse Demirkan, University of Surrey, UK

Microbial systems we host are currently under in-depth investigation as their involvement in almost all human biological systems are shown. Gut microbiome particularly is known to have systemic effects on multiple tissue and organs, ranging from neurodegeneration to tumor growth, response to cancer therapy and finally reproductive and maternal health. In this talk we will be focusing on gut and vaginal microbiome individuality and how they could relate to pregnancy related outcomes.

Dr Yevheniya Sharhorodska

Yevheniya is an actively practicing medical geneticist in reproductive health with expertise spanning over the past ten years. She holds a PhD in Medical Genetics; with the subject of her doctoral training on maternal effects determined by one carbon metabolism and vascular endothelial growth factor genes in foetal heart defect risk. She has been conducting her research activity at the Institute of Hereditary Pathology, National Academy of Medical Sciences, Lviv, and National Scientific Centre for Radiation Medicine of the National Academy of Medical Sciences, Kyiv, Ukraine. In addition, Yevheniya holds the position of Assistant Professor at the Danylo Halytsky Lviv National Medical University, where she teaches the undergraduate course in medical genetics.

While dissecting the healthcare issues of her patients Yevheniya began to expand her expertise and experience in an epidemiology, bioinformatics, genomics for practical implementation of this knowledge in reproductive health. Today, Yevheniya is Co-PI of the LUCAR (Lviv Ukrainian Cohort for Advancing Reproductive Health) study. She also works with other large-scale European datasets focussing on women's health, in particular, on recurrent pregnancy loss and its relationship to other phenotypes.

Ukrainian study of Recurrent pregnancy loss and UK biobank as genomic research resources

Dr Yevheniya Sharhorodska, University of Surrey, UK & Institute of Hereditary Pathology, Lviv, Ukraine

Women's reproductive health is the least systematically evaluated set of phenotypes in human genetics, contrary to its importance at individual level. Although most of the physical and mental conditions, that women experience, are not directly related to reproduction, the latter has large impact on the former. The prevalence of women's reproductive issues rapidly increases with older maternal age at conception and overall ageing of European populations. As a result, women turn to in-vitro fertilisation, a popularised industry that exacerbates certain reproductive health issues, with miscarriage being one of them.

Miscarriage is defined as the spontaneous loss of a pregnancy before the foetus reaches viability, typically occurring within the first 22 (24) weeks of gestation. A diagnosis of recurrent pregnancy loss (RPL) is made after the loss of two or more pregnancies before 22 (24) weeks of gestation. RPL can be caused by various factors, including genetic chromosomal abnormalities, hormonal imbalances, uterine abnormalities, autoimmune disorders, blood clotting disorders, and¬ certain infections. However, in 50% of patients with RPL the cause of disorder remains unclear. In this talk, I will discuss large-scale genetic investigations into women's reproductive health evaluated through pregnancy loss recurrence in the Ukrainian study and the UK Biobank.

Dr Natalia Pervjakova

Natalia is an accomplished professional with a strong background in molecular biomedicine and a proven track record in leading and managing complex projects. With a PhD in Molecular Biomedicine from the University of Tartu, Natalia has specialized knowledge in the field of genomic imprinting in complex traits.

During Natalia's PhD studies, she conducted groundbreaking research on genomic imprinting, exploring its role in complex traits. Under the guidance of esteemed researchers, Reedik Mägi, PhD, and Prof Andres Metspalu, MD, PhD, Natalia successfully completed their PhD thesis titled "Genomic imprinting in complex traits," making significant contributions to the field.

In addition to their doctoral degree, Natalia holds an International Master of Science degree in Biotechnology and Biomedicine from the same university. Her master's thesis focused on a genome-wide association studies in the Estonian population, supervised by Reedik Mägi, PhD, and Prof Andres Metspalu, MD, PhD.

The Estonian Biobank's MyGenome Portal: A comprehensive platform for return of results to over 200,000 biobank participants

Dr Natalia Pervjakova, Estonian Biobank, Estonia

The MyGenome Portal (MGP) of the Estonian Biobank is a research platform with multiple objectives: provide individual results to over 200,000 biobank participants; ensure transparency regarding the use of participants' data; improve public health and genomic literacy; serve as a platform for research; and support precision medicine initiatives.

The MGP is structured into three sections: personalized results, educational content, and studies. The portal incorporates dynamic consent that allows participants to choose specific categories for the return of results and which research projects they wish to join.

The personalized results section currently provides participants with reports on type 2 diabetes (T2D) and coronary artery disease (CAD) as well as pharmacogenomics, caffeine metabolism and ancestry at different resolutions. For T2D and CAD, the reports incorporate 10-year and life-time cumulative risk calculations based on polygenic risk scores, lifestyle factors and prevalent diseases. The pharmacogenomics report covers 26 medications with high evidence recommendations. The educational section complements the personalized results by providing respective background information. The interactive components and visual aids of the reports together with the educational section empower participants to gain a comprehensive understanding of the topics covered, facilitating informed decision-making regarding their health.

Continued ►

The dedicated research section serves as a digital solution for data collection. Among the first projects is an investigation into the perceived impact of genetic information reported through the portal. The research section will expand to include questionnaires, randomized clinical studies, and other research initiatives.

The development of the MGP involved consulting with expert groups and a qualitative study to assess user experience. Participants expressed satisfaction with the user-friendly interface and overall concept of the portal. They also expressed interest in receiving personalized reports and recommendations for a healthy lifestyle. Next, a preliminary launch is planned with 10,000 participants, with the full launch scheduled by the end of 2023.

Beyond facilitating the communication of personalized risk information, the MGP acts as a testbed for precision medicine initiatives. The study group within the MGP enables the creation of various sub-populations that represent the general population, allowing estimation of the impact on public health. The MGP empowers individuals, and promotes research, contributing to the ongoing transformation of healthcare.

Dr Muhammad Awais

Muhammad was lucky to be part of the research (together with colleagues Sara Atito and Josef Kittler) which resulted in the first state-of-the-art (SOTA) masked image modelling (MIM) approach for vision transformers using the simple principles of heavy masking and recovery of information without using human annotated labels. Proposed MIM outperformed all existing self-supervised learning (SSL) SOTA methods including joint embedding-based architectures at that time. It marked a milestone in computer vision to become the first method which used self-supervised pretraining (SSP) to outperform supervised pretraining (SP). According to leading Artificial Intelligence (AI) researcher and chief AI scientist of Meta, Yann LeCun the MIM has revolutionised the SSL.

Muhammad did his PhD in machine learning at Surrey. After his PhD he joined Imperial College London's startup, Cortexica, where he was machine learning lead. They applied machine learning for large scale retrieval problems. Muhammad came back to Surrey at the end of 2016 where he is currently a senior lecture in machine learning leading the efforts on

Continued ►

novel foundation models and generative AI.

The focus of my research is on core AI/ML/DL algorithms and their applicability to a wide range of application areas. My research interests include foundation models, un/self-supervised learning, cross/multi-modal learning, theoretical insights and understanding of deep learning, computer vision, NLP, medical image analysis, audio, retrieval, biometrics, security.

ChatGPT and Foundation models for health care and omics

Dr Muhammad Awais, AI institute, FEPS, University of Surrey, UK

The focus of Muhammad's presentation will be to introduce foundation models and generative AI we being developed at Surrey. The foundation models are the major source of transformation in AI. Muhammad will introduce some of the downstream tasks (application of foundation models) in medical imaging and healthcare. Hopefully he will then present some initial investigation into the use of multimodal foundation models they have developed on omics and medical imaging data.

Dr Joe Foster

Joe graduated from the European Bioinformatics Institute and the University of Cambridge in 2008 with a PhD in Bioinformatics, exploring how large-scale public mass spectrometry proteomics data could be QC'd, harmonised, and re-analysed for new, larger insights. Swiftly moving to industry, Joe has spent the last decade continuing the theme of generating, connecting, and leveraging large multi-omic datasets towards improving human health. Starting his career at Affymetrix, Joe worked on deploying microarray technology into the clinic for cancer diagnostics. Following cancer, Joe turned to the number one cause of human ill-health: food (or rather, lack of it!), working on population-scale genotyping programs for plants, animals and humans. Today, Joe is Director of Platforms at Lifebit, helping to scale precision medicine efforts with the power of federation through Lifebit's platform.

Scaling precision medicine through federation; how the sum is greater than the parts

Dr Joe Foster, Lifebit, UK

Data is foundational to all research activity, and in modern biomedical research, this data is multi-modal and diverse. Genomic data, for example, is increasing diagnostic yield and reducing time to treatment for patients. It's de-risking clinical trials and accelerating time to market for life-saving drugs and treatments. Our research communities' appetite for data is insatiable and when applied to machine learning and analyses like GWAS it is clear why there is increased power in training increasingly larger datasets.

The great news is there is a large amount of biomedical data out there to enable research. The bad news is that as much as 97% of it is inaccessible due to security, compliance, and technology issues. Even worse, the data that is available is often technically unusable, failing the "IR" of FAIR data principles or rendered unusable by research environments reserved for only the savviest “high-code” users. The solution to this issue is federation, whereby data can be found, accessed, and analysed without ever moving it. This principle of leaving the data where it belongs under the custodianship of its owner is solving security and compliance challenges worldwide and enabling a level of collaboration for sensitive biomedical data that was previously impossible. In this talk, we will discuss the different types of federation and some real-world examples of how when it is combined with an intuitive research environment, it can be used to scale precision medicine.

Continued ▶

POSTERS

FACULTY OF HEALTH AND MEDICAL SCIENCES

University of Surrey
Guildford, GU2 7XH, UK

surrey.ac.uk