

AI, Big Data, and Pregnancy Loss

Wednesday, the 12th of July 2023

INNOVATION FOR HEALTH BUILDING, STAG HILL CAMPUS, UNIVERSITY OF SURREY

Scientific programme committee:

Prof Inga Prokopenko, Dr Ayse Demirkan, Dr Adam Mahdi, Dr Yevheniya Sharhorodska

Report

Background

Identification of the predisposing genomic and other risk factors may help in determining the causes of pregnancy loss (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 combine with the newly launched two (GPU- and CPU-based) HPC clusters enabling big data modeling. 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.

Event themes

A synthesis of the main themes and issues discussed that emerged from the papers and discussions. The workshop brochure is available to view [here](#).

Women's reproductive health represents the least systematically dissected by scientists set of phenotypes, 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, that

exacerbates certain reproductive health issues, including **pregnancy loss (PL)**. PL is the spontaneous demise of a pregnancy before the foetus reaches viability. Miscarriage is an early PL, occurring at up to 20 weeks of gestation, and associated with excessive bleeding, infection, anxiety, depression, infertility and an increased lifetime risk of cardiovascular disease. A stillbirth (or late PL) is the death or loss of a baby at or after 20 weeks of pregnancy or during delivery.

The prevalence of **sporadic loss in a clinically recognized pregnancy** is approximately 15%, while the prevalence of RPL is about 1–3% in women from general population. In their lifetime, fewer than 5% of women will experience two consecutive miscarriages, and only 1% experience three or more. The majority (about 60%) of miscarriages occur when an embryo develops with an abnormal number of chromosomes or structure alterations. The abovementioned outcome occurs more commonly in women of higher reproductive age. **Known causes of PL** include parental genetic factors, anatomic abnormalities, antiphospholipid syndrome, endocrine disorders and thrombophilias. Factors that influence success in future pregnancy include parents' genetics, maternal age, number of prior losses, and genetic abnormalities of miscarried products. Additionally, **maternal risk factors** represented by high blood pressure, obesity, cigarette smoking, thrombosis and diabetes, may increase the risk of PL. However, a **putative diagnosis for PL causes is made and treated in only ~50%** of patients with PL and the exact mechanism that leads from to miscarriage is unknown.

PL remains a health care problem with no improvement for over 30 years. Many studies demonstrated **contribution of variety factors into PL susceptibility** 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.

During the workshop we explored the **ways to improve health care and outcomes for PL**, where Surrey could make a difference. **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 and expanded our effort on the implementation on modern multidisciplinary approaches to PL prevention using AI.

We discussed the **integration of expertise and methods from diverse disciplines:** epidemiology, statistics, data management, bioinformatics, human genomics and multiomics of reproductive health using AI for prevention of PL.

Professor Ranjit Akolekar, Medway Foetal and Maternal Medicine Centre, Medway NHS Foundation Trust, UK, highlighted different **risks factors of PL** including the most **recent translational research advances and evidence into maternity services** is essential to ensure that women and families receive the high-quality care they rightfully deserve and that we are committed to delivering. Dr Julia Zoellner, Institute for Women's Health, University College London, UK, discussed the clinical and genetic **prediction of gestational diabetes and its relationship with pregnancy outcomes**.

Dr Natalia Pervjakova, discussed about the Estonian Biobank's MyGenome Portal. While the biobank in Estonia covers about 20% of its adult population and has collected data and samples about/from their participants for the past 20 years, the main goal of providing health information to general population is becoming an important task for the scientists. The Biobank has now introduced a comprehensive platform for return of results to over 200,000 biobank participants, the trials of which have been very successful. Natalia discussed, how the genetic information would be useful in all areas of **health risk prediction** and, as the first author on the important paper on the genetics of Gestational Diabetes Mellitus (Hum Mol Genet, 2022), highlighted the **need for better prediction of pregnancy loss** to incorporate in the Estonian Biobank MyGenome portal and other similar applications.

Dr Mohammad Awais, People-Centred AI institute, University of Surrey, UK introduced the notion about **foundation AI models**, how these could be helpful and directly applied to big data collected about patients in Maternal and foetal medicine, including clinical records, ultrasound information, images etc. He connected this introduction with explanations about generative AI, used in GPT and similar to discuss the uses and pitfalls of interaction with such platforms. The discussion followed about the **needs to implement foundation models to develop predictive models for pregnancy outcomes**.

The workshop also featured a number of poster presentations by international researchers and live stream was performed from the University of Surrey.

Feedback among the event participants

"I was attracted by the opportunity to learn how modern data processing methods can help understand and potentially minimize the factors that cause pregnancy loss. I was also interested to hear the valuable opinion of experts on the current knowledge of the risk factors for reproductive losses."

"(The workshop offered) Opportunity to discuss our research with specialist who have wider and deeper expertise than we have (at our University)"

Dr Zhanna Balkhiarova, University of Surrey commented: *"The workshop on "AI, Big Data, and Pregnancy Loss" provided me with valuable insights that will significantly benefit my future work in the field of maternal-foetal health. During this workshop, I gained in-depth knowledge on critical aspects, such as clinical approaches to Recurrent Pregnancy Loss (RPL) and the application of Microbiome and Virome analyses in preventing pregnancy loss. Furthermore, the workshop shed light on a range of comprehensive analytical tools that I can integrate into my research endeavours. Notably, I learned about the utilisation of AI for processing omics data, the importance of computing and transcription regulatory elements (TREs), harnessing the capabilities of Chat GPT, and the significance of biobanking in advancing my research. These insights and tools will undoubtedly enhance my ability to make meaningful contributions in the field, ultimately leading to improved maternal-foetal health outcomes."*

Next steps – Outcomes

The workshop enabled participants to network with speakers and engage in discussions, fostered the development of collaborative projects and grant proposals in the interdisciplinary field of genetics, statistics, human genomics, multiomics and AI applied to life sciences and medicine, especially in the area of preventing pregnancy loss.

Based on the discussions during the workshop, we plan a collaboration with the Medway Foetal and Maternal Medicine Centre, Medway NHS Foundation Trust, UK. We will collaborate with Prof Akolekar to obtain the data from his database and analyse it for dissection of relationships between variety of potential risk factors of PL and pregnancy outcomes, focussing on stillbirth and comparing them to life births. We are already working on the ethics application preparation for the approval at the University of Surrey. Moreover, the AI Institute supported a new PhD student position for the research on pregnancy loss. The student has enrolled since the workshop and will begin to work on the topic pregnancy loss this September'2023. Additionally, the manuscript about causal effects of infection on pregnancy loss is in preparation for submission to a scientific journal. This project is led by Dr Sharhorodska and supervised by Prof Prokopenko through collaboration with our long-standing international partner, Institute of Hereditary Pathology, Ukraine. Dr Sharhorodska has submitted a postdoctoral EU MSCA grant on the subject investigating risk factors of pregnancy loss, if successful, she will join Surrey in a new role. Besides, the collaborators from the University of Ferrara also joined the workshop via online platform. They have joined a new collaboration with us now on the genetic causes of recurrent pregnancy loss and recurrent miscarriage and the relationship with nausea symptoms.

Overall, the outcomes of the workshop 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.

Acknowledgements

Sponsors, helpers, and anyone deserving particular praise.

This workshop was funded through an Institute of Advanced Studies Workshop Grant to Professor Inga Prokopenko. The workshop was also supported in part by the University of Surrey People-Centred AI institute. We would like to express our gratitude to Mirela Dumic and Louise Jones from IAS for all their support and advice throughout the process.

We are thankful to all speakers, who contributed with their valuable knowledge to the subject discussions. Special Thanks to Prof Ranjit Akolekar and Dr Julia Zoellner for their clinical insights and availability for the day.

Together with the event organising committee members, Dr Ayse Demirkan, Dr Adam Mahdi, and Dr Yevheniya Sharhorodska, Prof Prokopenko has led the preparation of the Workshop. We are grateful to the administrative support to Mrs Inna Trush – her help was invaluable for the preparation and smooth running of the day-long event. Big thanks to the members of the Section of Statistical MultiOmics for all types of help provided in organisation of the event.