# Project proposal

yasa@itu.dk

February 1, 2024

# 1 Project proposal

## 1.1 Problem formulation

### 1.1.1 Background

The need to close data gaps in women's health requires a reconsideration of traditional health metrics and studies (Burns et al., 2023). Sex-disaggregated data offer valuable insights into disease prevention and can lead to more effective interventions. This benefits not just women, but potentially results in more accurate diagnostics and treatment plans for everyone.

Diagnostic and treatment paradigms are often influenced by male-centric data, or suffer from a lack of sex-disaggregated data. This might lead to potential misdiagnoses, suboptimal treatments, and unequal health outcomes between men and women (FemTechnology, 2023).

### 1.1.2 Problem

Current medical AI models are trained on limited datasets that do not differentiate between biological sexes, potentially exacerbating gender disparities. Relying on medical AI without disclosing the demographic composition of the dataset or discussing its impact on the findings may result in biased algorithms and discriminatory outcomes. Recognising biological sex differences and accounting for confounding variables is crucial to ensuring fairness in medical AI (FemTechnology, 2023).

## 1.2 Method

I will adopt a methodology similar to the one outlined in (Birhane, 2022). In the context of the research papers presented at MICCAI 2023 (MICCAI), I will investigate how these papers define and address demographic information and societal needs, and whether they train and evaluate data by biological sex in their machine learning models.

I will likely, but not exclusively, refer to the following literature and use data from one or more of these sources:

- *The Values Encoded in Machine Learning Research* (Birhane, 2022)

- *Gall Bladder Cancer Detection from US Images with only Image Level Labels* (Basu, 2023)

- *Automated CT Lung Cancer Screening Workflow using 3D Camera* (Teixeira, 2023)

- *Detection-Free Pipeline for Cervical Cancer Screening of Whole Slide Images* (Cao, 2023)

- *Anatomy-Informed Data Augmentation for Enhanced Prostate Cancer Detection* (Kovacs, 2023)

Papers were chosen from the MICCAI 2023 portfolio (MICCAI) semi-randomly by searching for papers related to a non sex-specific disease, in this case, "cancer".

I reviewed each paper for demographic information in their datasets. For example, one paper (Basu, 2023) utilised a dataset containing 255 images from 218 patients. These patients were further classified into non-malignant (171 patients) and malignant (47 patients). However, the paper did not include demographic details such as age, geolocation, or biological sex. Another paper (Teixeira, 2023) used data from 62,420 patients from 16 different sites across North America, Asia, and Europe. However, they did not provide demographic information on the patients' geographic distribution or discuss the relationship

between their findings and geolocation. The biological sex of the patients was also omitted.

In an experiment, I evaluated whether papers centring sex-specific diseases included demographic information. A third paper (Cao, 2023) on cervical cancer, a female-specific disease, used a dataset containing 5384 cervical cytopathological WSIs[1] collected from "collaborating hospitals" (Cao, 2023). Despite using the term "women" in the introduction, there was no mention of demographic information or names of the hospitals involved in the data collection. Lastly, a fourth paper (Kovacs, 2023) on prostate cancer, a male-specific disease, used data from 774 consecutive bi-parametric prostate MRI examinations from patients. This paper, unlike the previous one, did not acknowledge the disease's sex-specific nature or provide any other demographic information about the patients.

## 1.3 What will be handed in

The following will be delivered as my project:

1. A project report

2. A GitHub repository containing the code used in my research

# References

Papanai A. Gupta M. Gupta P. Arora C. Basu, S. Gall bladder cancer detection from US images with only image level labels. *arXiv*, 2023. URL https://doi.org/10.1007/978-3-031-43907-0_20.

Kalluri P. Card D. Agnew W. Dotan R. Bao M. Birhane, A. The values encoded in machine learning research. *arXiv*, 2022. URL https://doi.org/10.48550/arXiv.2106.15590.

Delaney Burns, Tara Grabowsky, Emma Kemble, and Lucy Pérez. Closing the data gaps in women's health. *McKinsey Company, Life Sciences Practice*, 2023.

Fei M. Cai J. Liu L. Zhang L. Wang Q. Cao, M. Detection-free pipeline for cervical cancer screening of whole slide images. *Medical Image Computing and Computer Assisted Intervention – MICCAI 2023, Lecture Notes in Computer Science. Springer Nature Switzerland, Cham, pp. 243–252*, 2023. URL https://doi.org/10.1007/978-3-031-43987-2_24.

FemTechnology. The gender data health gap: Harnessing ai's transformative power to bridge the gender health data divide. 2023. URL https://femtechnology.org/wp-content/uploads/2019/07/Copy-of-Gender-Data-Health-Gap-PFG.pdf.

Netzer N. Baumgartner M. Eith C. Bounias D. Meinzer C. Jäger P.F. Zhang K.S. Floca R. Schrader A. Isensee F. Gnirs R. Görtz M. Schütz V. Stenzinger A. Hohenfellner M. Schlemmer H.-P. Wolf I. Bonekamp D. Maier-Hein K.H. Kovacs, B. Anatomy-informed data augmentation for enhanced prostate cancer detection. *Medical Image Computing and Computer Assisted Intervention – MICCAI 2023, Lecture Notes in Computer Science. Springer Nature Switzerland, Cham, pp. 531–540*, 2023. URL https://doi.org/10.1007/978-3-031-43990-2_50.

MICCAI. Medical Image Computing and Computer Assisted Intervention. 2023.

Singh V. Tamersoy B. Prokein A. Kapoor A. Teixeira, B. Automated ct lung cancer screening workflow using 3d camera. *Medical Image Computing and Computer Assisted Intervention – MICCAI 2023, Lecture Notes in Computer Science. Springer Nature Switzerland, Cham, pp. 423–431*, 2023. URL https://doi.org/10.1007/978-3-031-43990-2_40.

---

[1]WSI: whole slide imaging