Endometriosis Early Detection

Daniel Moshe | Adi Haber

Workshop Overview

Analyzing medical data found in UK Biobank using machine learning tools.

Research Question

Defining a research question regarding risk factors of a certain medical condition.

Creating a Cohort

Creating a cohort of patients relevant to the research question and cleaning it.

Extracting Features

Examining various medical studies and scientific articles to find features. Then extracting them from the UKB.

Creating a Model

Creating a machine learning model to get a prediction for the research question and repeating the process to improve it.

Our Research Question

What are the key risk factors associated with the development and diagnosis of Endometriosis?

What Is Endometriosis?

- A chronic medical condition where endometrial tissue grows outside the uterus and adheres to other organs, mainly in the pelvic region.
- Main symptoms include chronic pain, infertility, fatigue, and sometimes even anxiety and depression.
- Endometriosis is prevalent mostly in women of reproductive age (15-49). Researchers say 10% of this population is affected - about 190 million patients worldwide.

Reasons for Choosing this Subject

- Endometriosis is under-researched, with significant gaps in understanding its causes, risk factors, and optimal diagnostic methods.
- The diagnosis process is long and tedious.
- There is a lack of awareness to this illness, both from the general population and medical professionals.

The Problem

Diagnosis dificulty

- Endometriosis is very hard to diagnose.
- Adhesions are hard to see using imaging.
- Diagnosis time averages at 7 years.

Effects of delay

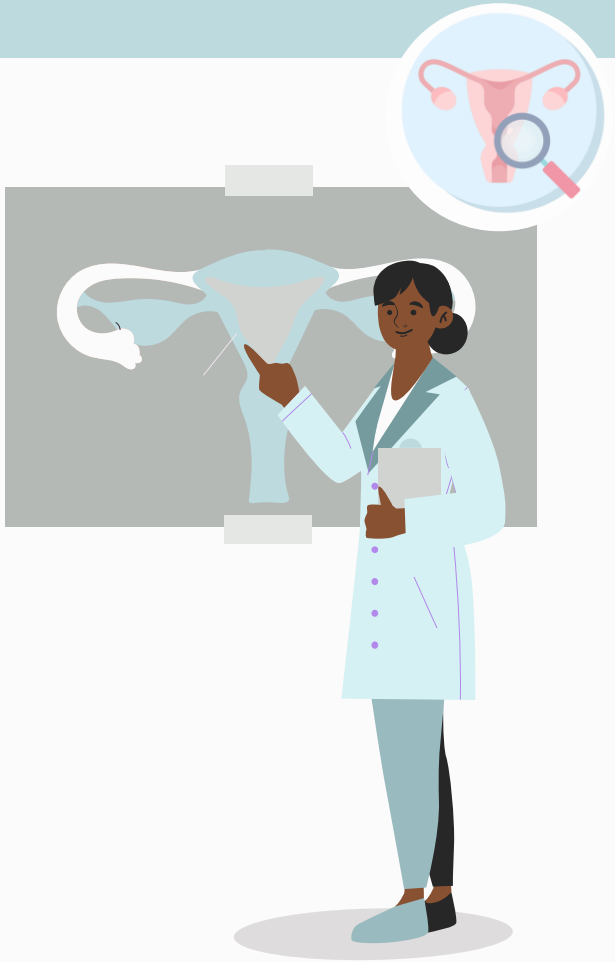
- Intensifies symptoms.
- Lowers quality of life.
- Causes incurable reproductive health challenges and infertility.

Current state

- Conventional diagnostic methods include invasive procedures.
- Doctors have subjective assessments.

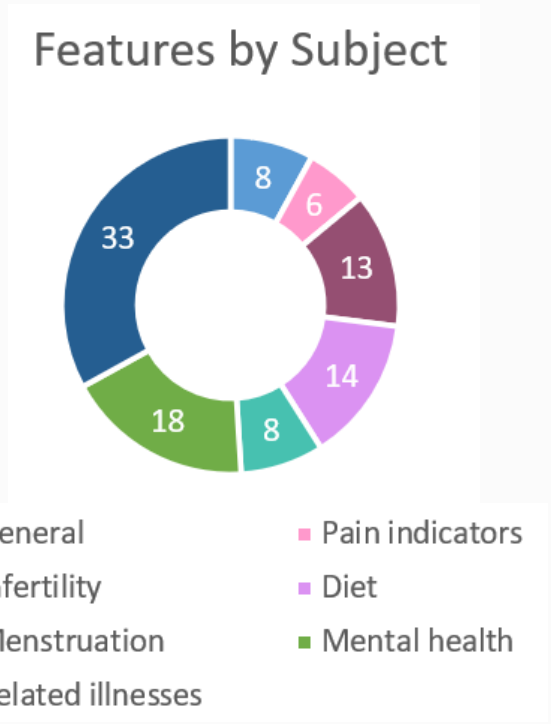
Proposed Solution

- Diagnosing Endometriosis by analysing UK Biobank patient data.
- Developing a machine-learning model for precise Endometriosis detection using these features extracted from the UK Biobank.

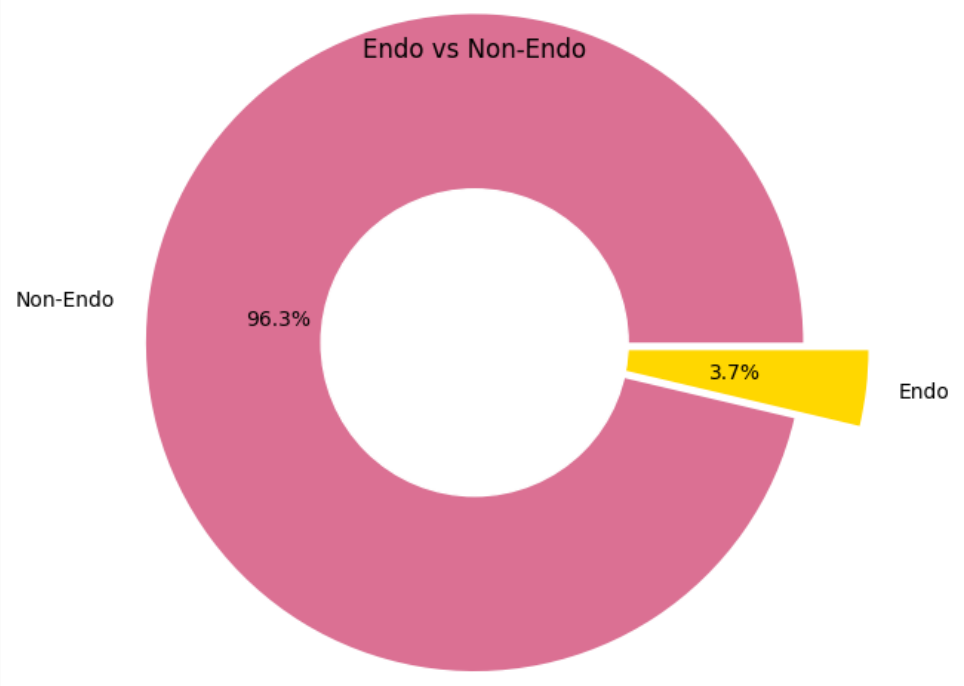


Extracting Initial Dataset

- We created a list of 100 features by feature id, from the UK Biobank Showcase.
- They include general data, pain indicators, infertility and pregnancy difficulties, diet, menstruation, mental health and related diseases.
- Coding a generic library for feature extraction.



Exploring our Data



~270,000

Female patients in the biobank

~10,170

Patients diagnosed with Endometriosis

Dilemmas and Challenges

Diagnosed Ratio

The dataset has about 4% diagnosed patients, while the diagnosis ratio in the general population ratio is 10%.

Train and Test Split

What should be the ratio of Endometriosis diagnosed patients and healthy patients in our datasets?

Diagnosis Age

The average diagnosis age in the dataset is 42, while the average woman will be diagnosed in her 30s. Will this affect the age as a feature?

Healthy Patients

Which of the non-positive patients should we choose for our train set? Exclude patients with other gynecological conditions?

Roadmap

01 Choosing and defining reaserch question

02 Research Edometriosis and find features

03 Code scripts for data extractions

04 Extract initial dataset

05 Explore and cleanse data

06 Feature engineering

07 Choosing model

08 Evaluation of the model

Thank you

Endometriosis Early Detection | Daniel Moshe, Adi Haber