Endometriosis Early Detection

Daniel Moshe | Adi Haber

Seminar 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.

Extracting Features

Examining various medical studies and scientific articles to find features.

Creating a Model

Creating a machine learning model to get a prediction for the research question.

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 area.
- These adhesions lead to inflammation and scar tissue forming on the affected organ, causing severe pain and infertility in some cases.
- Endometriosis is prevalent mostly in women of reproductive age, and researchers say 10% of this population globally is affected - about 190 million patients worldwide.

Reasons for Choosing this Subject

- Endometriosis is a condition that remains 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

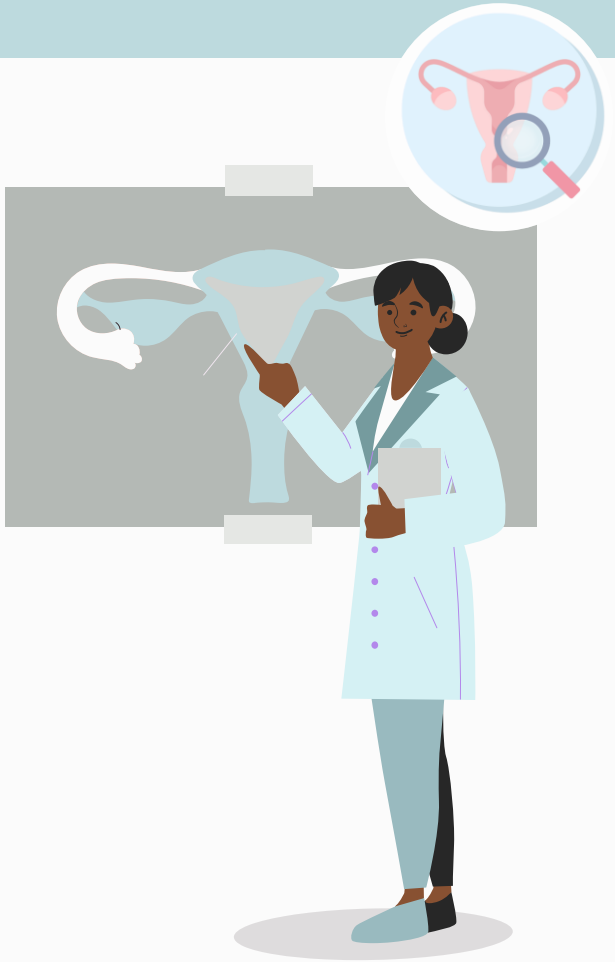
- Intensifies symptoms.
- Lowers quality of life.
- Causes incurable reproductive health challenges and infertility.

Current state

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

Proposed Solution

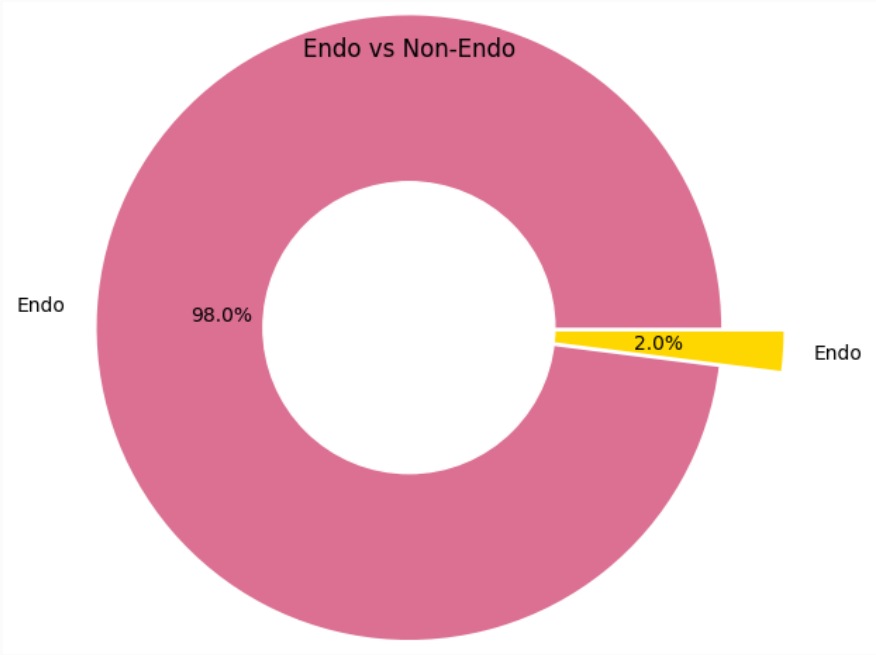
- This project aims to diagnose Endometriosis by analysing UK Biobank patient data.
- Gathering data on both Endometriosis and healthy patients and extract a set of features to detect Endometriosis.
- Our main objective is to develop the most effective machine-learning model for precise Endometriosis detection using these features.



Extracting Initial Dataset

- Choosing features based on published research and medical papers.
- Creating a list of features available in the UK Biobank by feature id, from the UK Biobank Showcase.
- 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 2% 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?

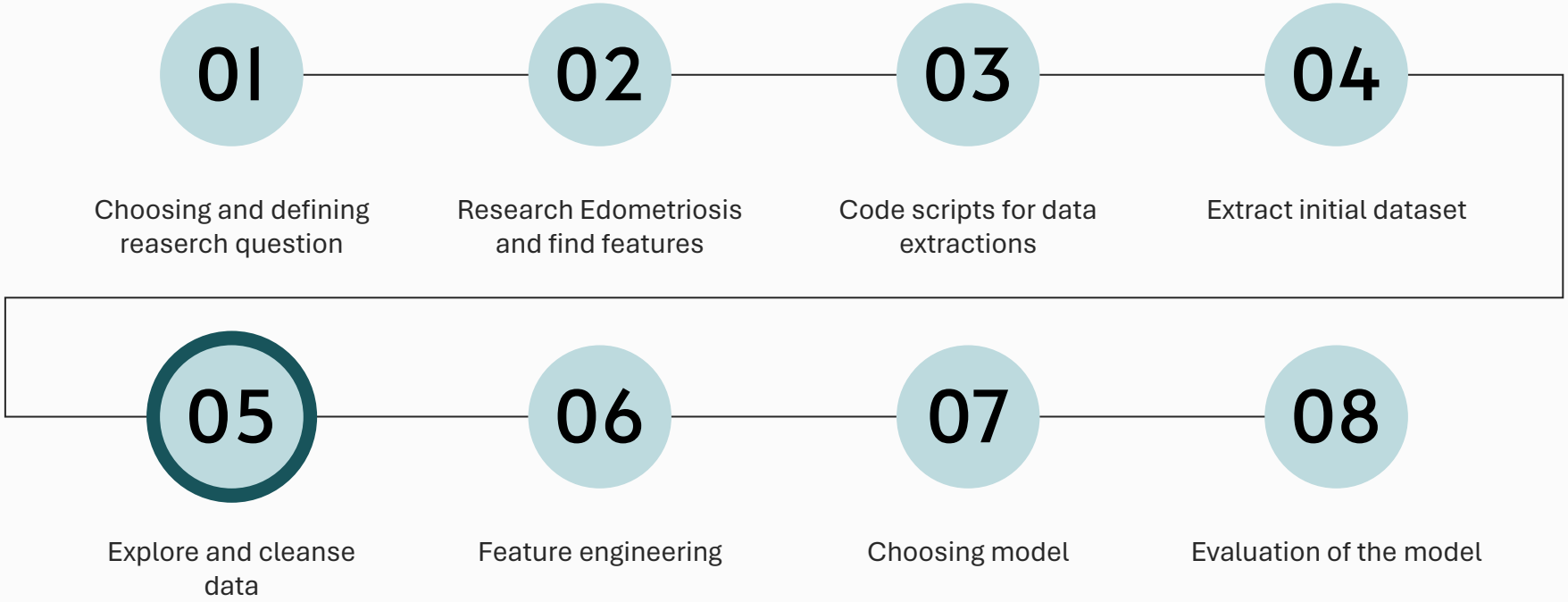
Diagnosis Age

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