Endometriosis Early Detection

Adi Haber | Daniel Moshe

Endometriosis 101

Endometriosis is a chronic condition where endometrial tissue grows outside the uterus, causing **pain** and **infertility**.

The Problem

Endometriosis is difficult to diagnose. Diagnosis takes an average of **7 years**. Current diagnosis methods include long **process of elimination** or **surgery**.
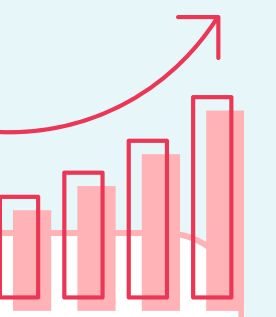
Our Solution

A machine learning model to identify common **risk factors** and **predict** endo, that will shorten diagnosis time.

The Data

The Patients:
- Female
- Reproductive age
- 10,000 diagnosed with endo
- 50% diagnosed in our cohort
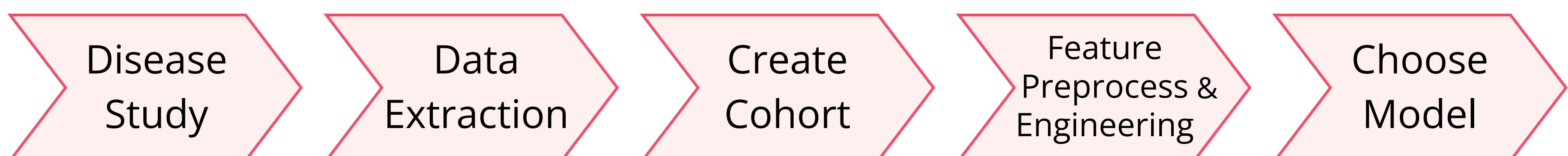
The Features:
- Pain
- Infertility
- Diet
- menstruation
- Mental health
- related conditions

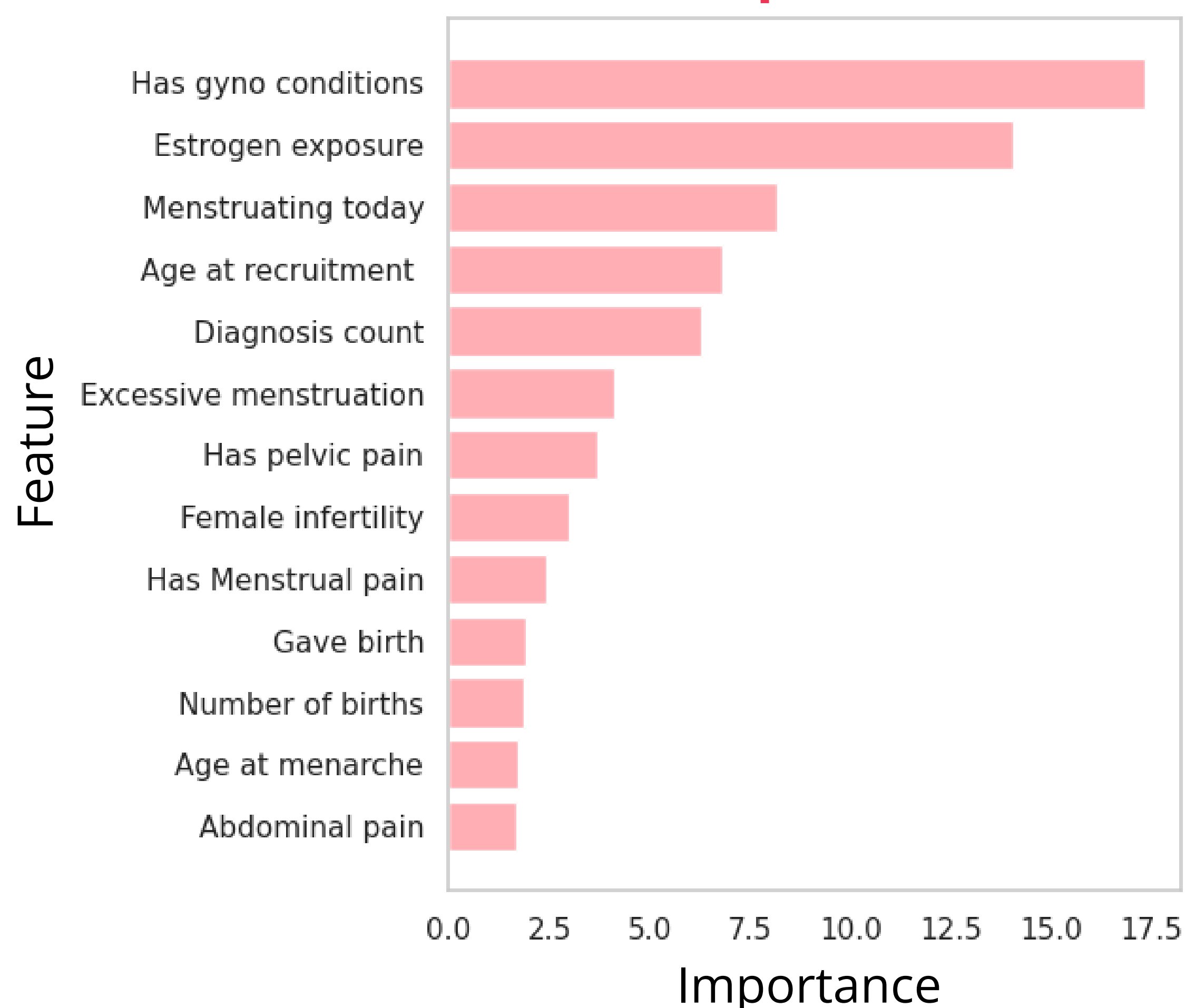**Feature Selection:** Chosen by known risk factors and various research studies.

The Model

CatBoost, a gradient boosting algorithm that efficiently handles categorical features.

The Process

Disease Study → Data Extraction → Create Cohort → Feature Preprocess & Engineering → Choose Model

Results

CatBoost Feature Importance

Feature importance bar chart (Importance 0.0–17.5):
- Has gyno conditions
- Estrogen exposure
- Menstruating today
- Age at recruitment
- Diagnosis count
- Excessive menstruation
- Has pelvic pain
- Female infertility
- Has Menstrual pain
- Gave birth
- Number of births
- Age at menarche
- Abdominal pain

Risk Factors

High Risk:
- Other gyno conditions
- Excessive menstruation
- IBS

Low Risk:
- Taking contraception pills
- High vegetable intake

Predictions

- Accuracy: 0.73
- F1 score: 0.72
- Precision: 0.74
- Recall 0.70

Confusion Matrix:
	Predicted 0	Predicted 1
Actual 0	2265	768
Actual 1	882	2155