Endometriosis early detection

Daniel Moshe, Adi Haber

Contents

Background

Endometriosis, a chronic inflammatory condition, primarily manifests through symptoms such as pain and infertility[1]. It occurs when tissue resembling the uterine lining grows outside the uterus, adhering to pelvic organs, and occasionally other areas of the body. This abnormal adhesion triggers inflammation and the formation of scar tissue, resulting in debilitating pain and, in some instances, infertility. Endometriosis predominantly affects women of reproductive age, with research suggesting that approximately 5-10% of this demographic, totalling around 180 million individuals globally, are affected[2].

Main known indicators of the endometriosis include:

- Pelvic pain and/or lower abdominal pain
- Painful menstrual cramps
- Abnormal menstrual bleeding pattern (either by amount or irregularity)
- Family history[3]
- Infertility[4]

Diagnosing endometriosis presents a challenge since adhesions are not always detectable through imaging techniques like ultrasound or MRI. Typically, a definitive diagnosis necessitates undergoing laparoscopic surgery[3] and a subsequent biopsy.

Problem Statement

Our primary objective revolves around the prompt identification of endometriosis. Remarkably, 60% of women dealing with endometriosis navigate consultations with three or more clinicians before receiving a diagnosis, leading to an average delay of seven years before definitive confirmation[5]. This prolonged delay intensifies symptoms, diminishes overall quality of life, and contributes to enduring reproductive health challenges. Conventional diagnostic methods, predominantly reliant on invasive procedures and subjective assessments, further compound the diagnostic process.

Proposed Solution

This project endeavors to aid in the detection of endometriosis by analyzing patient data, known risk factors, and symptoms to discern their correlation with endometriosis diagnosis. Our primary goal is to determine the most influential factors and identify the optimal machine learning model for accurate detection.

Introduction

Machine Learning

Machine learning, a subset of artificial intelligence, revolutionizes medical research by extracting insights from vast datasets to enhance diagnostic accuracy, treatment efficacy, patient outcomes, and notably, identifying risk factors.
Supervised learning algorithms, guided by labeled data, train models to predict outcomes or classify instances, offering valuable insights into disease detection and prognosis.
Unsupervised learning techniques uncover hidden patterns within unlabeled data, enabling researchers to identify novel disease subtypes or biomarkers.
Deep learning, a subset of machine learning, utilizes artificial neural networks with multiple layers to automatically extract complex features from raw data, paving the way for advanced image analysis, genomic sequencing, and more complex medical issues.
Through the integration of these machine learning paradigms, medical researchers unlock unprecedented opportunities to unravel the complexities of diseases, revolutionizing healthcare delivery.

Endometriosis

Endometriosis, a prevalent chronic gynecological condition reliant on estrogen, involves the presence of uterine endometrial tissue outside its normal cavity. This disorder manifests through superficial or deep pelvic peritoneal implants, adhesions, and ovarian cysts (endometriomas). Common symptoms encompass pelvic pain and infertility.

UK BioBank[6]

UK Biobank is a large-scale biomedical database and research resource, containing in-depth, de-identified genetic and health information from half a million UK participants. The database, which is regularly augmented with additional data, is globally accessible to approved researchers and scientists undertaking vital research into the most common and life-threatening diseases. UK Biobank's research resource is a major contributor to the advancement of modern medicine and treatment and has enabled several scientific discoveries that improve human health.

Related Works

Several research papers have applied machine learning techniques to predict endometriosis. These studies serve as a starting point for our research, allowing us to refine our unique research question and build upon existing knowledge in the field.

Study	Link	Date of Publication
Revisiting the Risk Factors for Endometriosis: A Machine Learning Approach	https://www.mdpi.com/1716284	Jul-2022
Diagnosis of Endometriosis Based on Comorbidities: A Machine Learning Approach	https://www.mdpi.com/2554716	Nov-2023
Machine learning algorithms as new screening approach for patients with endometriosis	https://doi.org/10.1038/s41598-021-04637-2	Jan-2022

Solution Description

Stages

In the initial phase of our research, we comprehensively explore the realm of endometriosis research, delving into the medical domain to gain a precise understanding of the parameters embedded within our dataset. We then embark on the process of data selection from the UK Biobank dataset, a vast repository containing diverse medical information, much of which may be irrelevant to our study on endometriosis. We meticulously curate this dataset, ensuring the inclusion of only pertinent data points and features. Subsequently, to comprehensively explore the realm of endometriosis research, it is essential to delve into the medical domain and gain a precise understanding of the parameters embedded within our dataset. By doing so, we can discern the significance of each parameter and optimize the dataset accordingly to serve as input for various machine learning algorithms. This optimization process enables us to harness familiar tools for scrutinizing our predictive models and developing a specialized model tailored to address the specific challenges posed by endometriosis. Ultimately, our goal is to assess and compare the results produced by various methodologies, with the aim of recommending the most comprehensive and precise approach for predicting endometriosis based on symptoms.

Extracting Data

In our endeavor to extract meaningful insights from the UK Biobank dataset for predicting endometriosis percentages, we encounter several challenges stemming from the dataset's vastness and heterogeneity. One primary issue is the abundance of irrelevant data that does not pertain to our research question, necessitating meticulous data selection. Additionally, a considerable portion of the records within the dataset contains missing values, further complicating our analysis. As we select appropriate features for our predictive models, we will confront the risk of inadvertently excluding records that possess missing values for other selected features, thereby potentially rendering our dataset incomplete. To mitigate this challenge, we will employ various data imputation techniques to address missing values, such as mean or median imputation, predictive imputation using machine learning algorithms, or deletion of records with missing values if appropriate. Furthermore, we will undertake rigorous data cleaning processes to ensure the integrity and reliability of our dataset, eliminating outliers and erroneous entries. Despite these challenges, our approach will aim to extract the most relevant and informative data from the entirety of the UK Biobank dataset, enabling us to develop robust predictive models for endometriosis percentages.

Tools and Technologies

In our research, we aim to leverage a variety of tools and technologies to tackle the challenge of predicting endometriosis percentages from tabular data, framing it as a binary classification problem. To begin, we will explore classical machine learning models such as logistic regression, decision trees, random forests, and support vector machines. These models provide a solid foundation for understanding the relationships between features and the target variable. Additionally, we plan to experiment with more complex models including gradient boosting machines (GBM) and ensemble methods to capture intricate patterns and interactions within the data. Furthermore, we will employ feature engineering techniques to extract meaningful insights and enhance model performance. This may involve creating new features, transforming existing ones, or selecting the most relevant features through techniques like recursive feature elimination or feature importance analysis. Finally, we intend to explore deep learning models utilizing neural networks (NN) to uncover deeper layers of abstraction and potentially capture nonlinear relationships present in the data. By employing this diverse array of tools and techniques, we aim to identify the most effective approach for accurately predicting endometriosis and ultimately contributing to advancements in medical research and diagnosis.

Anticipated Challenges

In our research, we confront several notable constraints that shape the scope and reliability of our findings. Firstly, while utilizing the UK Biobank dataset provides valuable insights, its demographic skew towards women averaging 50 years old presents a limitation. Given that our objective is to assist in diagnosing endometriosis in younger women, the dataset may not fully represent the nuances of the condition in this demographic. Moreover, a substantial proportion of cases labelled as 'has endometriosis' are self-diagnosed, comprising approximately 40% of our dataset. This introduces a layer of uncertainty regarding the accuracy of the diagnoses, as self-diagnoses may lack the rigor and precision of clinical assessments. Furthermore, the dynamic nature of endometriosis progression and treatment outcomes necessitates longitudinal data, which may be limited in our dataset. Despite these constraints, our research strives to navigate these complexities and contribute towards advancing the understanding and diagnosis of endometriosis.

Bibliography

[1] Wang, P.-H., Yang, S.-T., Chang, W.-H., Liu, C.-H., Lee, F.-K., & Lee, W.-L. (2022). Endometriosis: Part I. Basic concept. Taiwanese Journal of Obstetrics and Gynecology, 61(6), 927–934. https://doi.org/10.1016/j.tjog.2022.08.002

[2] Zondervan, K. T., Becker, C. M., Koga, K., Missmer, S. A., Taylor, R. N., & Viganò, P. (2018). Endometriosis. Nature reviews. Disease primers, 4(1), 9. https://doi.org/10.1038/s41572-018-0008-5

[3] Blass, I., Sahar, T., Shraibman, A., Ofer, D., Rappoport, N., & Linial, M. (2022). Revisiting the Risk Factors for Endometriosis: A Machine Learning Approach. Journal of Personalized Medicine, 12(7), 1114–1114. https://doi.org/10.3390/jpm12071114

[4] Tom Gunnar Tanbo, & Péter Fedorcsák. (2017). Endometriosis-associated infertility: aspects of pathophysiological mechanisms and treatment options. Acta Obstetricia et Gynecologica Scandinavica, 96(6), 659–667. https://doi.org/10.1111/aogs.13082

[5] Horne, A. W., & Missmer, S. A. (2022). Pathophysiology, diagnosis, and management of endometriosis. BMJ (Clinical research ed.), 379, e070750. https://doi.org/10.1136/bmj-2022-070750

[6] UK BioBank (2015). Ukbiobank.ac.uk. https://www.ukbiobank.ac.uk/learn-more-about-uk-biobank/about-us