

PROJECT TITLE: EARLY STAGE PCOS ANALYSIS and DETECTING RELATED COMORBIDITIES

PROJECT OVERVIEW:

Polycystic Ovary Syndrome(PCOS) is one of the most common disorder growing among women worldwide. Researchers have implemented various techniques to diagnose PCOS at an early stage. One of the primary diagnostic criteria includes examination of ovaries seen by ultrasound images in terms of number, size, and follicle distribution inside the ovary. This process includes manual tracing and follicle counting on the ultrasound images to decide PCOS.

As stated in a research paper by NIH (pubmed on 22 April 2024) that early detection of PCOS is crucial as it leads to long term effects on metabolic and reproductive health. PCOS also causes other comorbidities like, cardiac diseases, metabolic syndromes, resistance to insulin, infertility, and many more. According NHS diagnosis criteria we have three main symptoms. Any patient who's suffering from at least two symptoms are give proper PCOS treatment. Those symptoms are: irregular periods, high level of testosterone, and ovarian cysts. According to the NHS website: It's difficult to know exactly how many women have PCOS, but it's thought to be very common, affecting about 1 in every 5 women in the UK.

[Soni and Vashisht, 2018] highlighted symptoms and many risks associated with PCOS, treatment convenient for PCOS, and many Data mining methods that can be used to detect the PCOS with high rate of accuracy.

The time involving in clinical tests and scanning of ovary has become freight for PCOS patients. In order to counter this problem [Denny et al., 2019] proposed an system which will detect PCOS at an early stage.

While studying these overviews and research paper and keeping in mind I have included two research questions in my project. One is 'Early stage PCOS analysis' and another one is 'Detecting related comorbidities'. Diagnosis can generally be accomplished with a careful history, physical examination, and basic laboratory testing, with ultrasonography or other imaging. Hyperandrogenism can be diagnosed clinically by the presence of excessive acne, androgenic alopecia, or hirsutism (terminal hair in a male-pattern distribution).I have included two data sets, one is having clinical data and related health comorbidities and other is ultrasonography imaging(for detecting polycystic ovary). A polycystic ovary is defined as an ovary containing 12 or more follicles (or 25 or more follicles using new ultrasound technology) measuring 2 to 9 mm in diameter or an ovary that has a volume of greater than 10 mL on ultrasonography. A single ovary meeting either or both of these definitions is sufficient for diagnosis of polycystic ovaries.

RESEARCH QUESTION:

Which ML model will give best accuracy in detecting early stage PCOS and predicting related Comorbidities as well?

This project aims to develop a machine learning (ML)-based system to diagnose early-stage PCOS and detect associated comorbidities using clinical, biochemical, and imaging data. Two types of data needs to be collected. (1)Clinical Data which includes BMI and other hormonal data. (2)Biochemical Data which might be Ultrasound Images. Data Preprocessing is performed to standardize data, Extract Features, Label data, Finding missing values. Exploratory Data Analysis(EDA) is performed to Distribute the data. Also we can find correlation between the attributes which will be useful in detecting PCOS and also identifying any patterns. Then we will be applying ML model. Training and testing various ML algorithms (Random Forest, SVM, Neural Networks) and Optimizing hyperparameters for accuracy improvement. Identifying risks for diabetes, hypertension, and metabolic disorders. Using correlation analysis and predictive modelling. Then comes validation and deployment of results.

PROJECT TIMELINE:

TASK 1	TASK 2	TASK 3	TIMELINE
Literature Review(studying 2-3 papers selectively)	Choosing second dataset(for detecting relate comorbidities)	Preparing PDM plan	1 February - 9 February
Data Loading(code)	Looking into GitHub account	Distribution Analysis	11 February – 18 February
Data Cleaning	Missing Analysis	(Study Research papers as well)	19 February – 28 February
Data correlation Analysis	Train and test sets	Pipeline	1 March-15 March
Ethics preparation	Model implementation	Hyperparameter Tuning	16 March – 31 March

Report Writing	Report Writing	Cross Checking Code and Repository	1 April – 20 April
Revising Project	Preparing for Viva	Prepare for Viva	21 April – 12 My

DATA MANAGEMENT PLAN:

Dataset 1 is clinical data about the patient including BMI(Body Mask Index), Menstrual information and hormonal data. I have taken dataset from Kaggle.

<https://www.kaggle.com/datasets/ayamoheddine/pcos-dataset>

This data set is of tabular format of size 86KB . Dataset also include biochemical data which is used to predict related comorbidities like hypertension, mantal illness, thyroid. I have chosen Ultrasound image data of PCOS and Non PCOS patients.

https://figshare.com/articles/dataset/PCOS_Dataset/27682557?file=50407062

This dataset is in form of images categorised under affected and non affected attributes3,81,651KB. The Datasets I have chosen ae UH ethical and are publicly available.

REFERENCES:

Amsy Denny, Anita Raj, Ashi Ashok, C Maneesh Ram, and Remya George. Detection and prediction system for polycystic ovary syndrome (PCOS) using machine learning techniques. In TENCON 2019 - 2019 IEEE Region 10 Conference (TENCON).

Palvi Soni and Shevat Vashisht. Exploration on polycystic ovarian syndrome and data mining techniques. In 2018 3rd International Conference on Communication and Electronics Systems (ICCES).

<https://pubmed.ncbi.nlm.nih.gov/38779261/>

<https://pmc.ncbi.nlm.nih.gov/articles/PMC10137609/#abstract1>

<https://www.nhs.uk/conditions/polycystic-ovary-syndrome-pcos/>

<https://bjgp.org/content/70/694/e322>

