Proposal of Prediction Of Polycystic Ovary Syndrome (PCOS) Diagnosis Using Artificial Neural Network Algorithm.

By :
Zainab Ali Albashah MCS231021

Introduction: What Is Polycystic ovary syndrome (PCOS)

Polycystic ovary syndrome, also known as PCOS is a medical disorder characterized by excessive production of androgens, which are male sex hormones typically seen in women in modest quantities.

Polycystic ovarian syndrome, or PCOS, is characterized by the presence of several tiny cysts, which are fluid-filled sacs, in the ovaries. Nevertheless, several women diagnosed with this illness have no signs of the presence of cysts, while some women unaffected by the disorder do indeed develop cysts. Ovulation is the process in which a fully developed egg is discharged from an ovary. This occurs to facilitate fertilization by a male sperm. If the egg is not fertilized, it is expelled from the body during menstruation.

Background of the Problem:

Polycystic Ovary Syndrome (PCOS) is a prevalent endocrine disorder that affects a significant proportion of women of reproductive age. Characterized by symptoms such as irregular menstrual cycles, hirsutism, acne, and polycystic ovaries, PCOS is a complex condition with multifactorial etiology, including genetic, hormonal, and environmental factors. The disorder not only impacts reproductive health but is also associated with metabolic abnormalities, such as insulin resistance, obesity, and an increased risk of type 2 diabetes and cardiovascular diseases.

Statement of the Problem

The diagnosis is often made based on the patient's signs, symptoms, and physical examination. With the acquisition of knowledge and practical skills, nearly all medical professionals possess the ability to effectively diagnose PCOS. Within the field of medicine, the process of determining the nature of a medical condition, known as diagnosis.

Dealing with PCOS can be challenging and repetitive. Predicting PCOS based on numerous elements or symptoms is a complex issue with multiple layers. This is due to the fact that such behavior can result in incorrect assumptions and unforeseeable outcomes.

As a result of this issue, the healthcare business generates a dataset for the purpose of analyzing diseases.

UTM
UNIVERSITI TEKNOLOGI MALAYSIA

Research Questions:

1.What are the most common symptom of PCOS?

2.How accurately can a multiple regression model predict the presence of PCOS?

3.How can an Artificial Neural Network (ANN) be utilized to enhance the prediction of PCOS?

4.What insights can be gained from visualizing prediction outcomes using an interactive dashboard?

5.How can the findings from this research contribute to improving the diagnosis and management of PCOS in clinical practice?

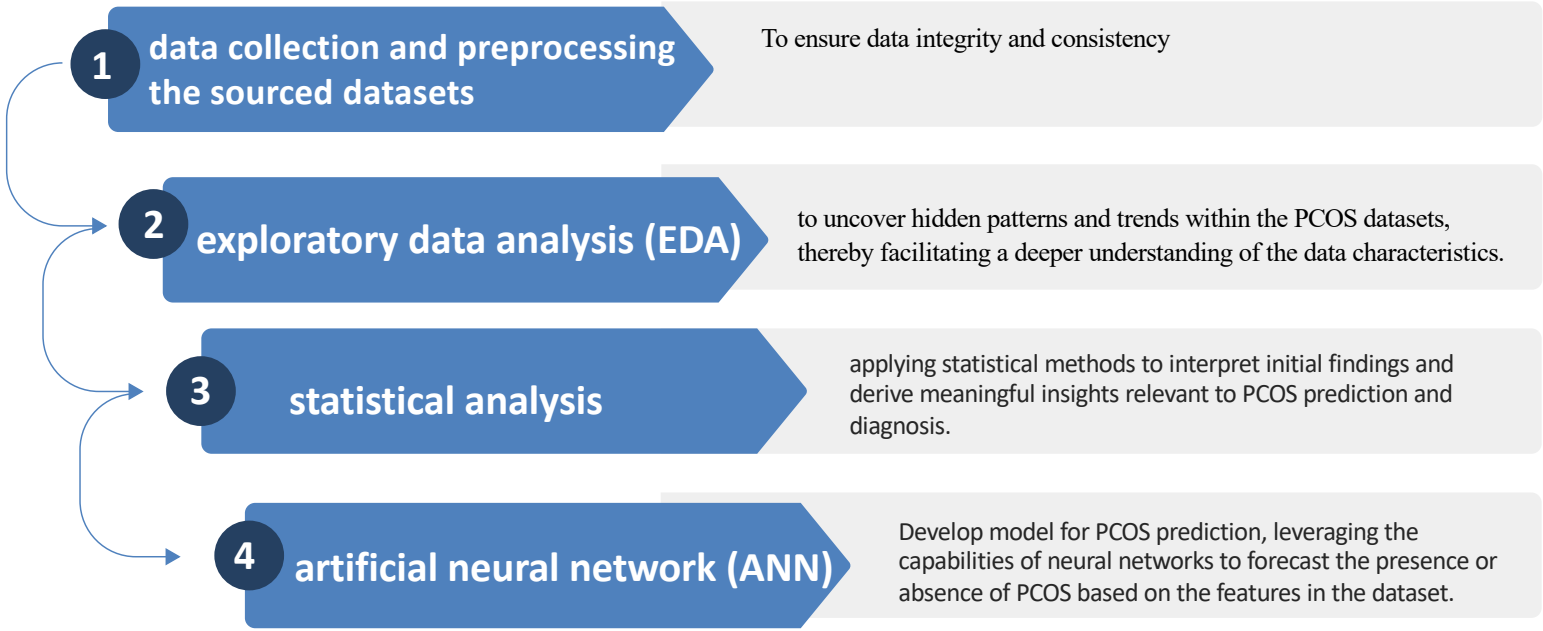
Objectives of the Research:

1.To determine the impact of factors on PCOS, a multiple regression model will be employed.

2.To develop an Artificial Neural Network (ANN) model for the PCOS dataset.

3.To create an interactive dashboard that visualizes and compares the prediction outcomes of a regression model and an Artificial Neural Network (ANN).

Scope of the Study:



Methodological Considerations in PCOS Research & Identifying the Research Gaps

PCOS, first described in 1935 by Stein and Leventhal, was initially characterized by ovarian cysts. Over time, its definition has expanded to include a range of symptoms and metabolic concerns. Globally, PCOS affects 5-10% of reproductive-age women (Dai et al., 2022).

In Western countries, extensive research explores genetic, hormonal, and environmental factors influencing PCOS. However, disparities exist in research and healthcare access globally, impacting diagnosis and management, especially in developing regions. Different ethnic groups show varying rates of PCOS, with metabolic issues posing additional challenges. Emotional impacts such as body image concerns and infertility can also contribute to social stigma and mental health issues.

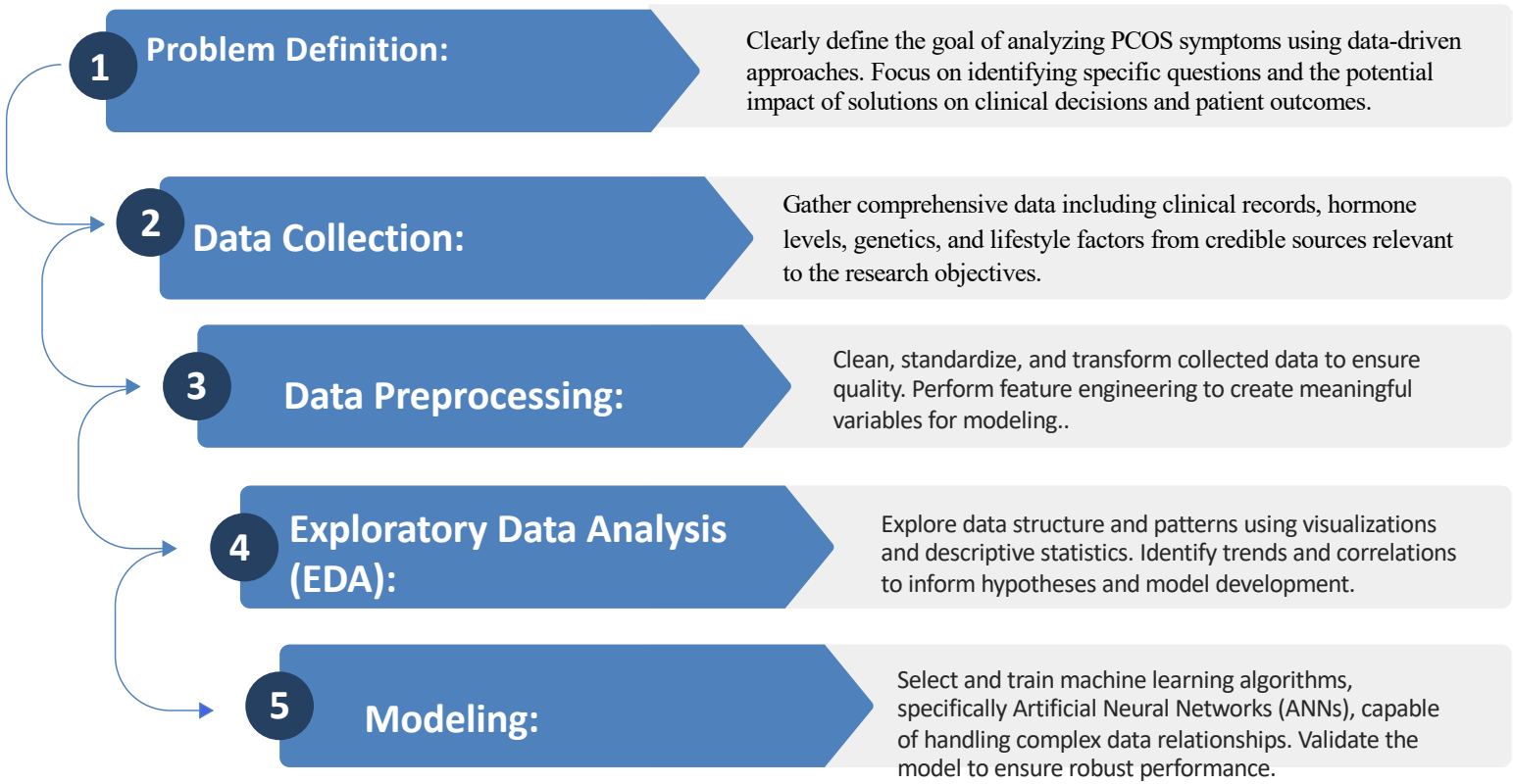
Literature Review/Problem Background
Historical Context and Global Perspectives

In PCOS research, methodologies vary widely, from clinical trials to population-based studies utilizing hormone profiling, ultrasound, and genetic analyses to identify PCOS markers. However, diverse study designs, sample characteristics, and diagnostic criteria pose challenges in synthesizing findings.

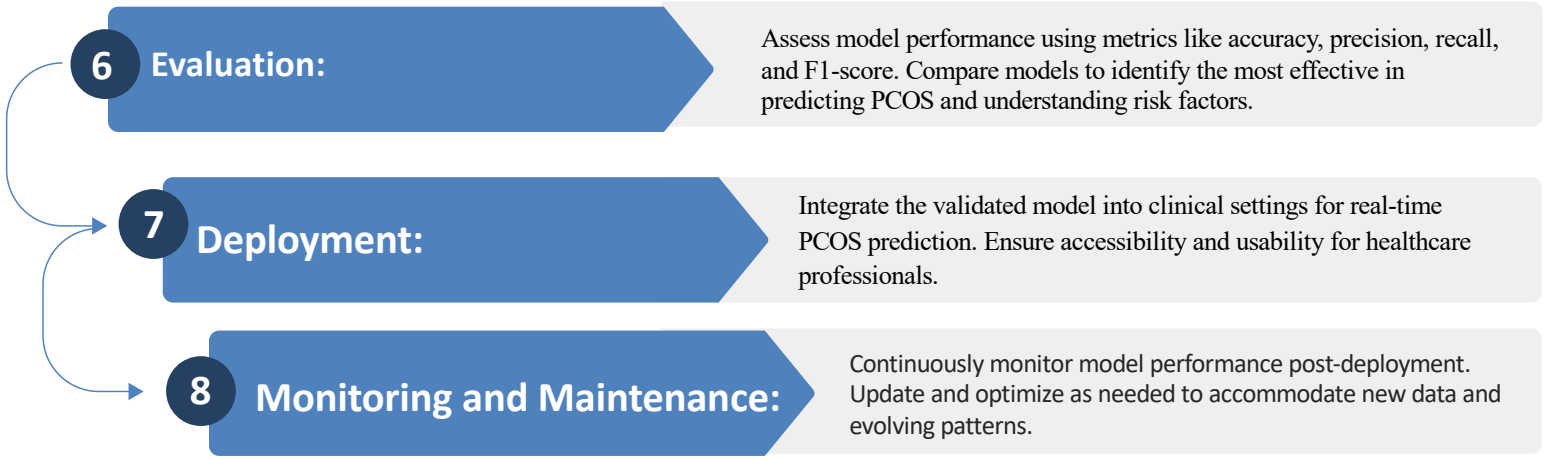
Some studies focus on biochemical aspects, while others emphasize clinical or metabolic impacts, highlighting the need for standardized diagnostic criteria

Identifying gaps in PCOS research is crucial. Long studies are lacking to track symptom progression and health risks over time, essential for evaluating management strategies. Psychosocial aspects, including anxiety, depression, and body image issues, are also underexplored despite their significant impact on patients. Culturally specific studies are needed to understand how PCOS manifests across populations and to develop culturally sensitive healthcare strategies.

Data Science Project Life Cycle



Data Science Project Life Cycle



Data Sources and Collection Methods

The data for this study was sourced from Kaggle and they involve a combination of clinical records, patient surveys, and publicly available datasets on PCOS. The datasets included variables such as age, BMI, insulin levels, androgen levels, menstrual cycle regularity, and the presence of polycystic ovaries.

	Sl. No	Patient File No.	PCOS (Y/N)	Age (yrs)	Weight (Kg)	Height(Cm)	BMI	Blood Group	Pulse rate(bpm)	RR (breaths/min)	...	Pimples(Y/N)	Fast food (Y/N)	Reg.Exercise(Y/N)	BP _Systolic (mmHg)	BP _Diastolic (mmHg)
0	1	1	0	28	44.6	152.0	19.300000	15	78	22	...	0	1.0	0	110	80
1	2	2	0	36	65.0	161.5	24.921163	15	74	20	...	0	0.0	0	120	70
2	3	3	1	33	68.8	165.0	25.270891	11	72	18	...	1	1.0	0	120	80
3	4	4	0	37	65.0	148.0	29.674945	13	72	20	...	0	0.0	0	120	70
4	5	5	0	25	52.0	161.0	20.060954	11	72	18	...	0	0.0	0	120	80

Data Pre-processing

Before analysis, the collected two datasets merging into one data underwent several pre-processing steps to ensure quality and consistency:

Data Cleaning:
Categorical Encoding:

0	Sl. No	541 non-null	int64
1	Patient File No.	541 non-null	int64
2	PCOS (Y/N)	541 non-null	int64
3	Age (yrs)	541 non-null	int64
4	Weight (Kg)	541 non-null	float64
5	Height(Cm)	541 non-null	float64
6	BMI	541 non-null	float64
7	Blood Group	541 non-null	int64
8	Pulse rate(bpm)	541 non-null	int64
9	RR (breaths/min)	541 non-null	int64
10	Hb(g/dl)	541 non-null	float64
11	Cycle(R/I)	541 non-null	int64
12	Cycle length(days)	541 non-null	int64
13	Marraige Status (Yrs)	540 non-null	float64
14	Pregnant(Y/N)	541 non-null	int64
15	No. of aborptions	541 non-null	int64
16	I beta-HCG(mIU/mL)	541 non-null	float64
17	II beta-HCG(mIU/mL)	541 non-null	object
18	FSH(mIU/mL)	541 non-null	float64
19	LH(mIU/mL)	541 non-null	float64
20	FSH/LH	541 non-null	float64
21	Hip(inch)	541 non-null	int64
22	Waist(inch)	541 non-null	int64
23	Waist:Hip Ratio	541 non-null	float64

Exploratory Data Analysis (EDA)

Statistic Analyzing
Blood Group indications

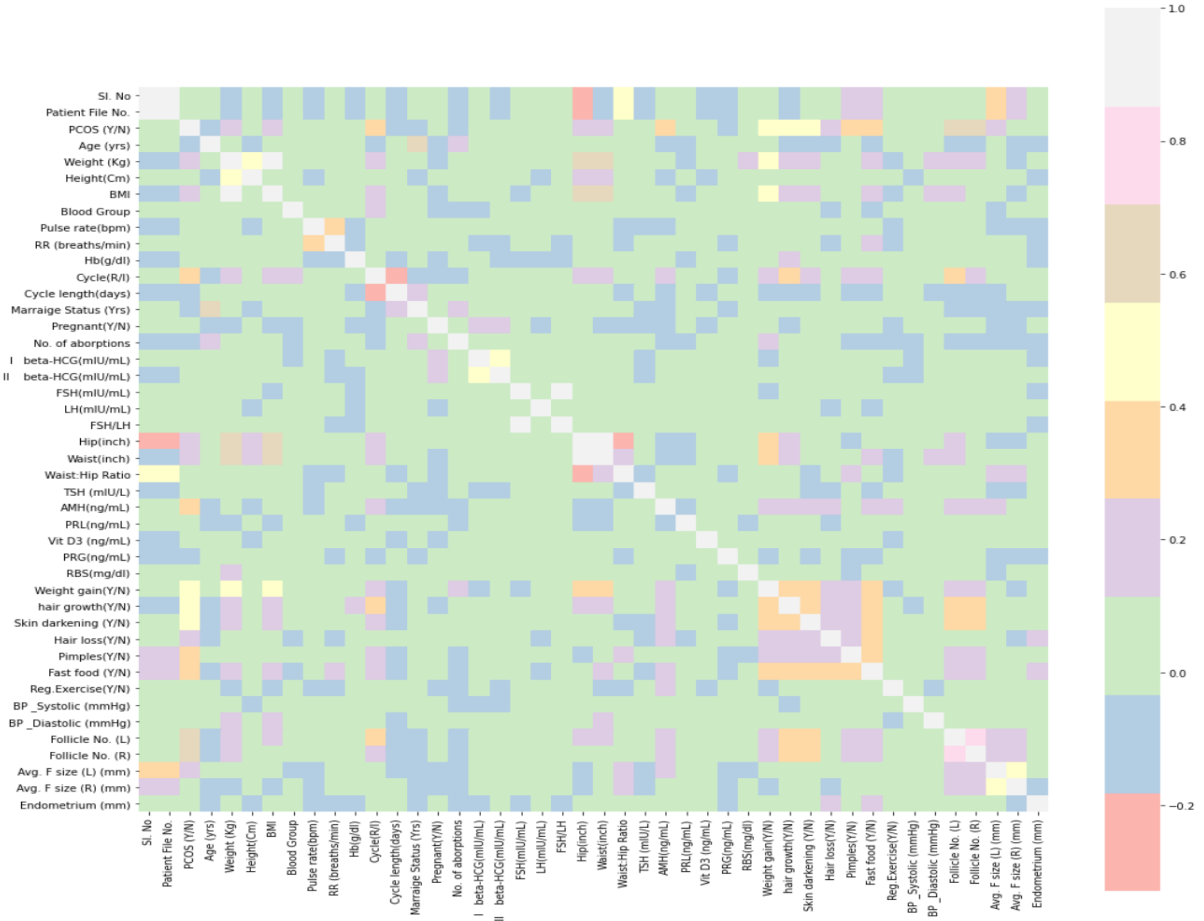
A+ = 11
A- = 12
B+ = 13
B- = 14
O+ =15
O- = 16
AB+ =17
AB- = 18

	Sl. No	Patient File No.	PCOS (Y/N)	Age (yrs)	Weight (Kg)	Height(Cm)	BMI	Blood Group	Pulse rate(bpm)	RR (breaths/min)	...	Pimples(Y/N)	Fast food (Y/N)
count	541.000000	541.000000	541.000000	541.000000	541.000000	541.000000	541.000000	541.000000	541.000000	541.000000	...	541.000000	541.000000
mean	271.000000	271.000000	0.327172	31.430684	59.637153	156.484835	24.311285	13.802218	73.247689	19.243993	...	0.489834	0.515712
std	156.317519	156.317519	0.469615	5.411006	11.028287	6.033545	4.056399	1.840812	4.430285	1.688629	...	0.500359	0.500216
min	1.000000	1.000000	0.000000	20.000000	31.000000	137.000000	12.417882	11.000000	13.000000	16.000000	...	0.000000	0.000000
25%	136.000000	136.000000	0.000000	28.000000	52.000000	152.000000	21.641274	13.000000	72.000000	18.000000	...	0.000000	0.000000
50%	271.000000	271.000000	0.000000	31.000000	59.000000	156.000000	24.238227	14.000000	72.000000	18.000000	...	0.000000	1.000000
75%	406.000000	406.000000	1.000000	35.000000	65.000000	160.000000	26.634958	15.000000	74.000000	20.000000	...	1.000000	1.000000
max	541.000000	541.000000	1.000000	48.000000	108.000000	180.000000	38.900000	18.000000	82.000000	28.000000	...	1.000000	1.000000

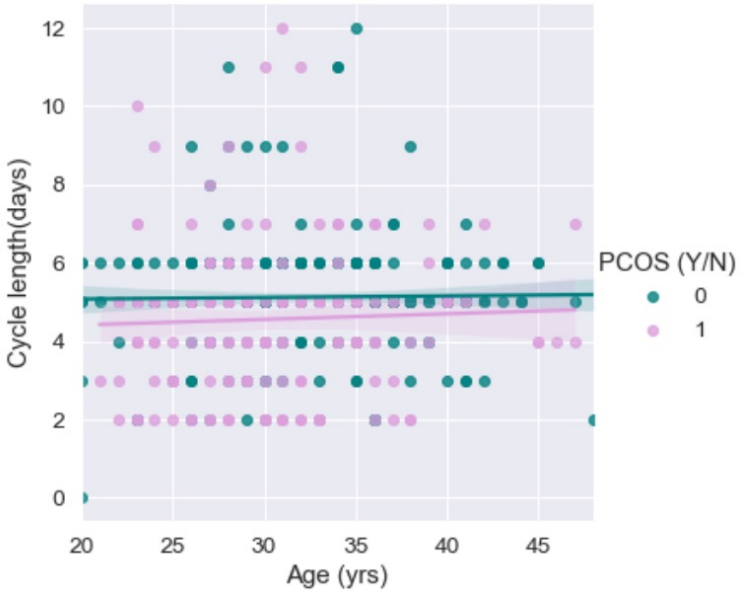
Exploratory Data Analysis (EDA)

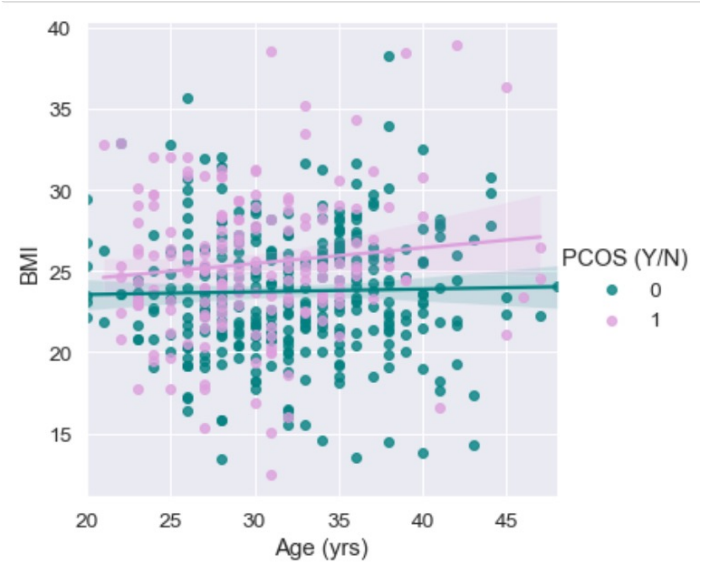
Examining a correlation matrix of all the features

at the right of the figure, we can see there is a high correlation between weight gain, hair growth (in undesired spots), skin darkening, hair loss, pimples, and fast food, at the right of figure, at the middle we can see there is a correlation between hip, waist size, and waist-hip ratio when waist size increase the hip size increase also.



Exploratory Data Analysis (EDA)





shutterstock.com · 2177064579