PREDICTION OF POLYCYSTIC OVARY SYNDROME (PCOS) DIAGNOSIS USING ARTIFICIAL NEURAL NETWORK ALGORITHM

ZAINAB ALI ALBASHAH

ABSTRACT

The primary focus of this research is to investigate the factors influencing Polycystic Ovary Syndrome (PCOS) and its implications both globally and locally, given its substantial medical and socio-economic consequences. This critical health issue necessitates detailed research due to the limited number of existing studies and incomplete data. The study aims to understand the impact of various factors on PCOS, facilitating the development of predictive models and diagnostic tools to support effective decision-making in healthcare. By employing advanced analytical methods and comprehensive data collection, this research seeks to fill gaps in the current understanding of PCOS. The literature review situates the study within both historical and contemporary research contexts, highlighting the need for region-specific insights. The methodology chapter outlines the data science framework, detailing data sources, collection methods, and preprocessing steps to ensure robust analysis. Exploratory data analysis (EDA) utilizes visualizations and statistical techniques to examine the PCOS dataset, perform feature engineering, and develop predictive models for assessing PCOS risk. The discussion chapter presents findings from various models, including linear regression and Artificial Neural Networks (ANNs), and highlights their implications for PCOS diagnosis and management. This study aims to provide valuable insights for healthcare stakeholders, enhancing strategies for managing PCOS effectively under varying conditions. The ultimate goal is to contribute to improved healthcare approaches, benefiting both medical practice and policy-making in the management of PCOS.

CHAPTER 1

INTRODUCTION

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.

Occasionally, a woman may experience insufficient production of the necessary hormones for ovulation. If ovulation fails to occur, the ovaries may develop several tiny cysts. These cysts produce hormones known as androgens. Women diagnosed with Polycystic Ovary Syndrome (PCOS) frequently have elevated levels of androgens. This can exacerbate issues with a woman's menstrual cycle. Furthermore, it can induce a multitude of symptoms associated with Polycystic Ovary Syndrome (PCOS). Pharmaceutical intervention is commonly employed for the management of PCOS. While it does not possess the ability to treat PCOS, it does aid in diminishing symptoms and mitigating some health complications.

The precise etiology of PCOS remains uncertain. Insulin resistance is prevalent in numerous women with PCOS. This indicates that the body has a reduced ability to effectively utilize insulin. Elevated insulin levels accumulate in the body and can lead to increased androgen levels. Obesity can exacerbate PCOS symptoms by elevating insulin levels.

PCOS can potentially have a genetic component. Polycystic ovary syndrome (PCOS) is frequently observed in siblings or in mothers and daughters. If your mother or sibling has PCOS, there is a higher likelihood that you may also have it. If you have insulin resistance or are obese, you are more likely to having it.

1.2 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.

The traditional approach to diagnosing PCOS involves a combination of clinical evaluation, including the assessment of symptoms and physical examination, and laboratory tests to measure hormone levels. Imaging studies, such as ultrasound, are often used to identify polycystic ovaries. Despite the availability of these diagnostic tools, accurately diagnosing PCOS remains challenging due to the heterogeneity of its presentation and the overlap of its symptoms with other medical conditions.

In recent years, the healthcare industry has seen a significant increase in the generation and availability of data related to patient health. This data encompasses electronic medical records, laboratory results, imaging studies, and patient demographics. Big data analytics has emerged as a powerful tool to process and analyze this vast amount of information, providing insights that can enhance clinical decision-making and improve patient outcomes.

The application of big data in healthcare offers the potential to revolutionize the diagnosis and management of conditions like PCOS. By leveraging advanced data analytics techniques, it is possible to identify patterns and correlations that may not be apparent through traditional methods. This can lead to more accurate predictions, personalized treatment plans, and ultimately, better health outcomes for patients.

However, the integration of big data into clinical practice is not without challenges. Issues such as data quality, interoperability, and the need for specialized expertise in data analysis must be addressed. Furthermore, there is a need for robust predictive models that can handle the complexity of PCOS and provide reliable diagnostic support.

The background of this problem highlights the need for innovative approaches to improve the prediction and diagnosis of PCOS. By utilizing big data and exploratory analysis, this project aims to identify significant risk factors and underlying patterns that can enhance our understanding of PCOS and support the development of more effective diagnostic tools.

1.3 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. The healthcare industry generates large data for research on diseases, including information about patients, hospital resources, electronic medical

records, disease diagnosis, and medical equipment. Big data is an essential asset that needs to be processed and analyzed in order to extract knowledge. This knowledge can then be used to enhance decision-making and save expenses. Only relying on human intelligence is insufficient for accurate diagnosis. The diagnosing process will involve numerous challenges.

For instance, factors such as low precision in obtaining outcomes, inadequate expertise, and performance that is contingent upon the duration and complexity of knowledge enhancement. Thus, the main issues addressed in this project involve the prediction of PCOS by taking into account significant risk variables and studying the underlying patterns required for PCOS prediction using exploratory analysis.

1.4 Research Questions

1. What are the most common symptoms 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?

1.5 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).

1.6 Scope of the Study

This study focuses on the comprehensive analysis of datasets related to Polycystic Ovary Syndrome (PCOS). To ensure data integrity and consistency, the primary objectives include data collection and preprocessing the sourced datasets. Next, we conduct exploratory data analysis (EDA) to uncover hidden patterns and trends within the PCOS datasets, thereby facilitating a deeper understanding of the data characteristics.

Moreover, statistical analysis involves applying statistical methods to interpret initial findings and derive meaningful insights relevant to PCOS prediction and diagnosis. Lastly, machine learning model development We are developing an artificial neural network (ANN) model specifically tailored for PCOS prediction, leveraging the capabilities of neural networks to forecast the presence or absence of PCOS based on the features in the dataset.