International Journal of Research in Engineering and Science (IJRES)
ISSN (Online): 2320-9364, ISSN (Print): 2320-9356
www.ijres.org Volume 11 Issue 1 ǁ January 2023 ǁ PP. 165-170

Detection of Polycystic Ovary Syndrome (PCOS)Using Machine Learning Techniques

Dr. Jyothi R Associate Professor
Computer Science and Engineering Department
Global Academy of Technology Bangalore, India

1. Shivani H C
Computer Science and Engineering Department
Global Academy of Technology Bangalore, India

2. Vidyashree
Computer Science and Engineering Department
Global Academy of Technology Bangalore, India

3. Yashaswi R
Computer Science and Engineering Department
Global Academy of Technology Bangalore, India

4. Sumanth R
Computer Science and Engineering Department
Global Academy of Technology Bangalore, India

Abstract— *The current world ladies populace is generally impacted by preterm early terminations, fruitlessness, and an ovulation and so on. It has been shown that polycystic ovarian disease (PCOS), a syndrome common in women of reproductive age, has a substantial role in the aetiology of infertility. PCOS affects a large number of women of reproductive age. It is an endocrine condition distinguished by changes in female chemical levels and uncommon male chemical output. This illness results in ovarian failure, which increases the likelihood of failed labour and fruitlessness. PCOS symptoms include weight gain, an erratic monthly cycle, abnormal production of male hormones, skin irritation, and hirsuteness. Analyzing PCOS is highly challenging due to the variety of side effects and the presence of a varied number of connected gynecological disorders. For PCOS patients, the time and money spent on many clinical tests and ovary inspections has become a burden. To address this issue, this study proposes a model for the early diagnosis and prediction of PCOS on ideal and modest but promising clinical and metabolic indicators that act as an early warning sign for this ailment. Classification of PCOS with the list of capabilities changed with Head Part Examination is finished utilizing different machine learning methods like Irregular Woods classifier, Machine Learning and CNN (Convolution Neural Network) in Python IDLE.*

Keywords— Machine Learning, PCOS, Detection, and Classification

---

Date of Submission: 02-01-2023 Date of acceptance: 14-01-2023

I. INTRODUCTION

Preterm abortions, infertility, an ovulation, and other issues have a significant impact on the world's female population today. It has been noted that the condition known as polycystic ovarian syndrome (PCOS), which impacts reproductive-age women, has a significant role in the causes of infertility. PCOS affects more than five million women globally who are of reproductive age. It is an endocrine condition defined by alterations in the levels of female hormones and aberrant male hormone synthesis. Ovarian dysfunction brought on by this illness increases the risk of infertility and miscarriage. Obesity, an irregular menstrual cycle, an excess production of the male hormone, acne, and hirsute are all signs of PCOS. Due to the wide range of symptoms and the existence of numerous related factors, PCOS is very challenging to diagnose. One of the most frequent hormonal illnesses among women of reproductive age is polycystic ovarian syndrome (PCOS). This is a complex endocrine illness that increases the risk of infertility, ovulation problems, cardiovascular disease, type 2 diabetes, obesity, and other

issues. PCOS is a widespread disorder that affects 12-21% of women of reproductive age, with 70% of those affected going untreated.

Controlled medications and lifestyle modifications can help manage PCOS to some extent. This covers anti-androgens, birth control pills,diabetes, fertility, and scanning techniques including an ultrasound scan. The capacity of the ovary to ovulate is improved by reducing the level of male hormone when such treatments are unsuccessful, then invasive therapeutic methods such surgical drilling of the ovaries are then performed.

The diagnosis of PCOS is complicated and clinically unclear despite the existence of diagnostic criteria. Due to the complexity of the condition's symptoms and their wide range of severity, relatively littlestudy has been done on the diagnosis of PCOS. To give an example, it has been observed that PCOS has a range of heterogeneous presentations that differ based on age and ethnicity and can be suggestive of a range of additional co-morbidities. Although physicians are aware of the most prevalent co-morbidities, including obesity, diabetes, insulin resistance, and depression, they are less likely to make the connection between PCOS and disorders includingsleep problems, liver failure, endometrial cancer, gestational diabetes,and anxiety.

Furthermore, a lot of medical professionals overlook PCOS. According to a community-based survey, PCOS affects 18% of women, with 70% of cases going untreated. Additionally, recent evidence indicates that PCOS underdiagnosis is worsened, particularly in teens. Puberty also brings on the initial signs of PCOS, including acne, hormonal imbalances, the development of cysts, and the existence of numerous follicles.

Therefore, at that age, doctors are less worried about these symptoms Other issues with PCOS diagnosis have been brought up as some physicians expand the PCOS diagnostic criteria to include polycystic ovaries [1]. In order to lessen and minimize human error in PCOS diagnosis, the aim of this project is to build a high-performing diagnosticmodel using machine learning (ML). By incorporating ML into PCOS diagnosis, we can assist women who have the condition since early detection can help with early treatment of ailments, infertility problems, and involvement in an appropriate lifestyle to avoid weight gain and metabolic problems.

II. LITERATURE SURVEY

Deep Neural Network image segmentation for detecting polycystic ovarian disease. 2019 is the year of publication. This paper explains the proposed methodology for detecting PCOD disease. The proposed methodology will yield accurate results. The benefit here is that the Classification result will be able to classify the disease based on its severity, which can be high, medium, or low.

Machine Learning Methods for Polycystic Ovary Syndrome Detection (PCOS). The year of publishing is 2022. In this work, a high-performing model to forecast PCOS's existence or absence was developed using a labelled dataset. This has the advantage that the validation parameters have been assessed and the diagnosis has been forecasted using seven classifiers.

Detection of Polycystic Ovary Syndrome Using Machine Learning Techniques. The year of publishing is 2021. In this work, we examined the application of classification algorithms to recognize polycystic ovary syndrome patients. The advantage in this case is that their examination of the data showed that although the classifiers' recall performance was subpar, their accuracy was rather good. Generally speaking, the majority of the research in this area built a diagnostic model using the same processing and machine learning methodologies. The majority of study authors tested their conclusions on a limited sample size and omitted some medical details. But in this work, we suggest a PCOS diagnosis model based on elements that are scientifically valid. Additionally, despite being tested on several individuals, our model produced excellent results. Finally, based on physician recommendations, we chose precision as our primary outcome. This section discusses the suggested process for creating a PCOS diagnosis model. The dataset that was applied is described. The preprocessing procedures are then discussed, followed by a discussion of feature selection, model training, and model evaluation.

III. METHODOLOGY

The creation of a suitable machine learning model-based diagnostic tool for PCOS requires a comparison of the performance of several current algorithms in our data set. The preparation of the model, which gives the research its framework, is the most crucial phase. With the use of a workflow diagram and the procedures required in creating an acceptable model and fine-tuning it to provide the optimal outcome are described below. Additionally, it is necessary to discuss the useful tools and accessible platforms utilized for system development. The next section discusses both aspects.

I. Data Collection

This first step is to gather the relevant data from KAGGLE Analyze it and see if the information gathered is appropriate for our project.

II. Pre-processing The following step in pre-processing is to remove null values. Later, all categorical textual data is converted into binary numerical data to aid in prediction. It's a statistical technique for evenly spreading out the number of instances in your collection. It functions by exploring innovative instances from minority situations that have already occurred.
III. Feature Selection Making a choice feature that are Extremely Randomized Trees or Extended tree classifier ensemble machine learning technique is used to choose the characteristics that are crucial for PCOS illness prediction. It is analogous to classifiers from random forests. When compared to random forest, it differs in the way decision trees are constructed in the forest. Based on the values of feature importance, features are classified as best or relevant. The Extra Trees approach use a training dataset to generate a large number of unpruned decision trees. Predictions in regression are created by averaging the decision tree output, but predictions in classification are made by majority voting.

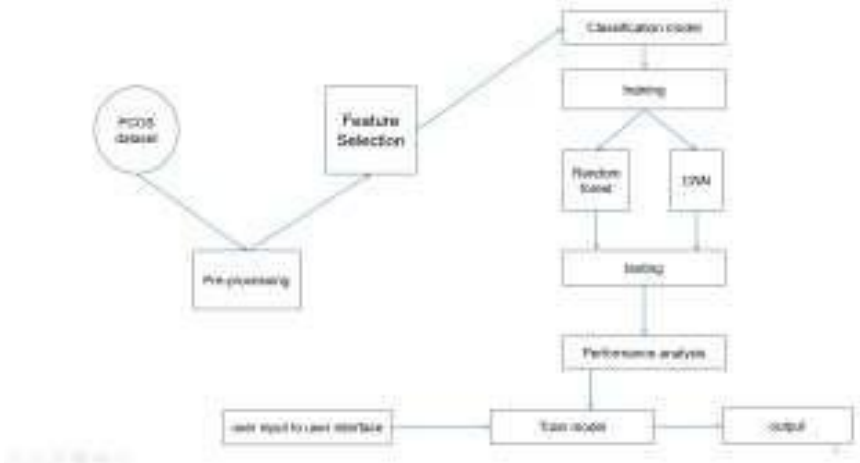
IV. Applying Algorithms
A. Random Forest Algorithm Random Forest Numerous single decision trees that make up Random Forest work together to form an ensemble. The class with the most votes will be the class that the random forest's trees predict, and that class will be the model's output. Any single constituent model will perform worse than an enormous number of relatively uncorrelated models (trees) acting as a single committee. The use of random forests has the benefits of quicker training and more accurate output prediction.

B. CNN (Convolution Neural Network)
Machine learning includes convolutional neural networks, sometimes known as convnets or CNNs. Convolutional neural networks, often known as convnets or CNNs, are a kind of machine learning. It is one of several artificial neural network models that are utilized for a variety of activities and data sets. A CNN is a type of network design used in deep learning algorithms for applications such as image recognition and pixel data processing. Primarily used for image processing.

V. Connection Part
Web Application Flask which is a python open-source software is used to build a web application, it is just a micro-framework it does not provide any database or form validation. HTML and CSS are used with flask which renders html pages and displays the content on the web page. CSS provides specifications for placement of html elements on the web page. Users can give various inputs like sugar level, specific gravity, albumin etc. After taking the inputs through the form, values from the html elements are taken by the flask functions and data frame is created out of inputs, which is given as a input to the trained model and based on the output received different web page is shown which tells whether the patient is having a disease or not.

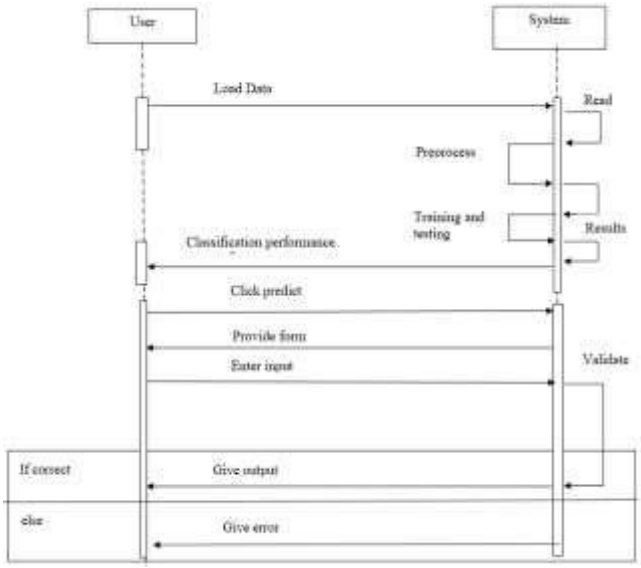
Design and Development Process

1. System Architecture

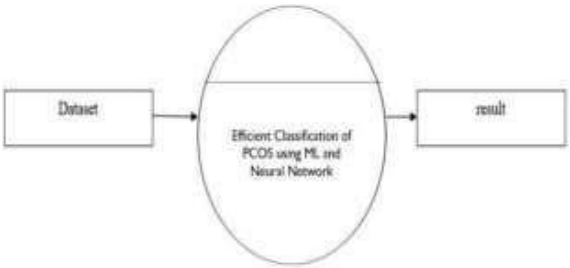


Prior to using algorithms, the data is pre-processed and features are chosen. Two algorithms served as the foundation for our comparison. Neural networks and Random Forest are taken into consideration while assessing performance.

2. Sequence Diagram:

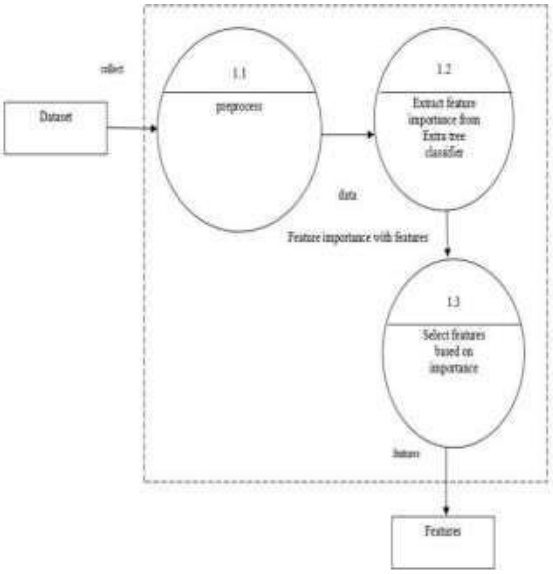


3. Data Flow Diagrams
Level 0:



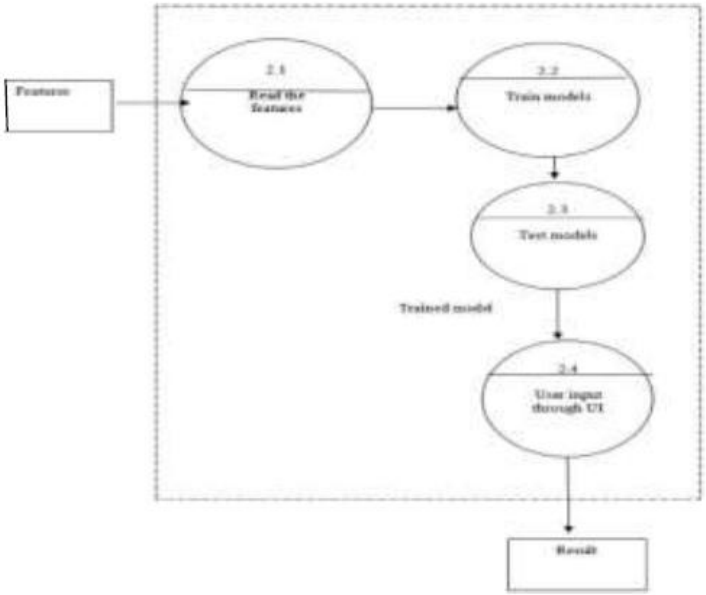
Results are produced once the dataset is fed into the machine learning and neural network models.

Level 1:



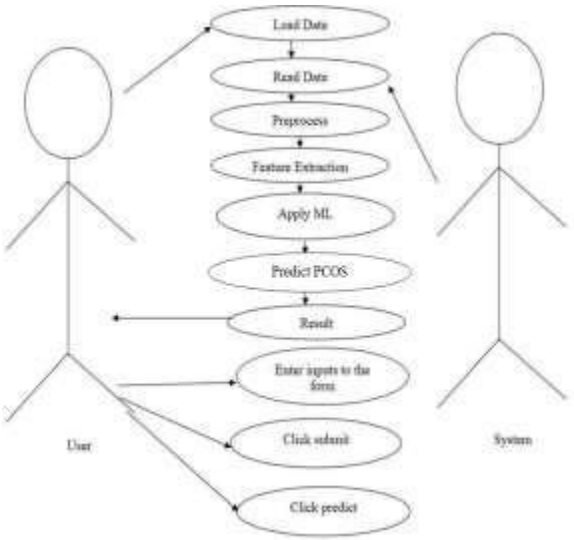
The dataset is initially gathered from several sources before being pre-processed to eliminate null values and analyze the data. The additional tree classifier is used to extract important characteristics.Important features with a high feature importance are chosen as topor important features based on their feature importance.

Level 2:



The dataset is initially gathered from several sources before being pre-processed to eliminate null values and analyses the data. The additional tree classifier is used to extract important characteristics.
Important features with a high feature importance are chosen as top or important features based on their feature importance.

Use Case Diagram:



The model loads and reads the dataset. To eliminate null values, thedata is pre-processed. Extra tree classifier is used for feature selection and extraction. The existence of renal disease is predictedusing ML models, and results are obtained. The user inputs are received from a form, and when the user clicks submit, the data is transmitted to the system, and when the user hits predict, the resultis discarded in the project of this PCOS project.

IV. RESULTS AND DISCUSSION

After comparing the results of the literature papers according to theanalysis contained in the writing, it was assumed that the proposed model used elements that were dependent on healing and had a high level of accuracy, both of which can still be improved through testing. Following all of the inputs given and read by the system, returns apositive or negative result based on the user's inputs.

V. CONCLUSIONS

Many women of reproductive age suffer from the endocrine condition known as polycystic ovary syndrome (PCOS). Ovulation and infertility may result from this. The clinical and metabolic data that serve as biomarkers for the illness are included in the diagnostic criteria. Based on a limited set of possible markers, we developed asystem that can identify PCOS. Using the data readily accessible, webuild a feature vector as part of our technique. Using feature selection algorithms, we will select the best features for predicting PCOS (POLYCYSTIC OVARIAN SYNDROME). The Decision Tree andCNN (Convolutional Neural Network) models will be analyzed andcompared. This system aids in the diagnosis of PCOS and enables thepatient to begin treatment as soon as possible before anything worseoccurs.

REFERENCES

[1]. Image Segmentation for detecting Polycystic Ovarian Disease using Deep Neural Networks Palvi Soni1*, Sheveta Vashisht2 Research Paper Vol. -7, Issue-3, March 2019

[2]. Detection of Polycystic Ovary Syndrome (PCOS) Using MachineLearning Algorithms Dana Hdaib, Dana Hdaib, Hiam Alquran, Wan Azani Mustafa, Wan Azani Mustafa, Ahmed Alkhayyat 5th International Conference on Engineering Technology and its applications 2022- (5thIICETA2022)

[3]. Automated Detection of Polycystic Ovary Syndrome Using Machine Learning Techniques Yasmine A. Abu Adla, Dalia G. Raydan, Mohammad-Zafer J. Charaf, Roua A. Saad, Jad Nasreddine, Mohammad O. Diab Conference Paper October 2021

[4]. Improved Sampling and Feature Selection Selection to supportExtreme Gradient Boosting For PCOS Diagnosis Muhammad Sakib Khan Inan, Rubaith E. Ulfath, Fahim Irfan Alam, Fateha Khanam Bappee, Rizwan Hasan 2021

[5]. Automated Screening of Polycystic Ovary Syndrome using Machine Learning Techniques Palak Mehrotra,Jyotirmoy Chatterjee, ChandanChakraborty, BiswanthGhoshdastidar,Sudarsh an Ghoshdastidar G D Institute for Fertility Research 2011