IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 2, August 2023

Analysis of PCOS using Machine Learning

Chandru A S[1] and Dr. Seetharam K[2]
Assistant Professor, Department of Information Science & Engg[1]
Professor, Department of CSE[2]
National Institute of Engineering North Campus (NIE North), Mysuru, Karnataka, India[1]
R.L Jalappa Institute of Technology, Doddaballapura, Karnataka, India[2]
chandru@nieit.ac.in and drseetharamkk@gmail.com

Abstract: *Polycystic Ovary Syndrome (PCOS) is a common endocrine disorder affecting women of reproductive age. Early detection and diagnosis of PCOS is crucial for the timely management of symptoms and prevention of long-term health complications. In this research paper, we propose an Android application for the detection of PCOS using Support Vector Machine (SVM) and data mining techniques. The proposed application will collect various physiological and demographic data of the user, such as age, height, weight, blood pressure, and menstrual cycle history. This data will be used to train an SVM model that will predict the likelihood of the user having PCOS. The application will also use data mining techniques to identify any patterns and trends in the collected data. To evaluate the performance of the proposed PCOS detection application, we will conduct a study using a sample population of women with and without PCOS. The collected data will be used to train and test the SVM model, and the accuracy, precision, recall, and F1-score of the model will be calculated. The proposed PCOS detection application will provide a non-invasive, cost-effective, and user-friendly tool for the early detection of PCOS. This application has the potential to improve the diagnosis and management of PCOS and ultimately improve the health outcomes of women with PCOS*

Keywords: Polycystic Ovary Syndrome

I. INTRODUCTION

Polycystic Ovary Syndrome (PCOS) is a common endocrine disorder affecting women of reproductive age. It is estimated that up to 20% of women worldwide may have PCOS, making it one of the most common endocrine disorders in women. PCOS is characterized by a combination of symptoms, including irregular menstrual cycles, insulin resistance, and hyperandrogenism, which can lead to infertility, weight gain, and long-term health complications such as type 2 diabetes and cardiovascular disease. Despite the high prevalence of PCOS, it remains underdiagnosed and undertreated. One of the challenges of PCOS diagnosis is the lack of a definitive diagnostic test, and diagnosis often relies on the presence of a combination of symptoms. This can lead to delays in diagnosis and treatment, which can impact the long-term health outcomes of women with PCOS. In recent years, there has been growing interest in the use of machine learning and data mining techniques for the early detection and diagnosis of PCOS. These techniques can analyze large datasets and identify patterns and trends that may not be apparent to human analysis. By leveraging these techniques, it may be possible to develop more accurate and efficient diagnostic tools for PCOS. In this research paper, we propose an Android application for the detection of PCOS using Support Vector Machine (SVM) and data mining techniques. The application has the potential to provide a non-invasive, cost-effective, and user-friendly tool for the early detection of PCOS, which can ultimately improve the health outcomes of women with PCOS

II. LITERATURE SURVEY

In recent years, there has been an increasing interest in the use of machine learning and data mining techniques for the early detection and diagnosis of PCOS. Several studies have explored the use of these techniques to develop more accurate and efficient diagnostic tools for PCOS.

One study by Saeedi et al. (2020) proposed a PCOS diagnosis system based on a fuzzy expert system and support vector machine algorithm. The system was able to achieve an accuracy of 92.5% in the diagnosis of PCOS. Another

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 2, August 2023

study by Raju and Krishnan (2019) used data mining techniques to analyze a large dataset of PCOS patients and identified several risk factors associated with the disorder, including age, body mass index, and serum testosterone levels. In a study by Kaur et al. (2018), a machine learning-based approach was used to develop a PCOS prediction model using a dataset of physiological and clinical parameters. The model was able to achieve an accuracy of 86.1% in predicting the presence of PCOS. In a more recent study by Sharma et al. (2021), an Android application was developed for the prediction of PCOS using a machine learning-based approach. The application was able to achieve an accuracy of 86.7% in the prediction of PCOS. These studies demonstrate the potential of machine learning and data mining techniques for the early detection and diagnosis of PCOS. However, there is still a need for more research in this area to develop more accurate and efficient diagnostic tools for PCOS. The proposed Android application using SVM and data mining techniques can contribute to this growing body of research and provide a valuable tool for the early detection of PCOS.

2.1 PROBLEM DEFINITION

Polycystic Ovary Syndrome (PCOS) is a common endocrine disorder affecting women of reproductive age. PCOS is characterized by a combination of symptoms, including irregular menstrual cycles, insulin resistance, and hyperandrogenism. It can lead to infertility, weight gain, and long-term health complications such as type 2 diabetes and cardiovascular disease. Despite its high prevalence, PCOS remains underdiagnosed and undertreated, and there is a lack of a definitive diagnostic test for PCOS. Diagnosis often relies on the presence of a combination of symptoms, leading to delays in diagnosis and treatment, which can impact the long-term health outcomes of women with PCOS.
The problem, therefore, is the need for a more accurate and efficient diagnostic tool for PCOS that can improve the early detection and management of the disorder. The proposed Android application using Support Vector Machine (SVM) and data mining techniques can address this problem by providing a non-invasive, cost-effective, and user-friendly tool for the early detection of PCOS. The application will collect various physiological and demographic data of the user, such as age, height, weight, blood pressure, and menstrual cycle history, and use this data to predict the likelihood of the user having PCOS. By providing an efficient and accurate tool for the early detection of PCOS, this application can improve the long-term health outcomes of women with PCOS.

2.2 OBJECTIVES

- Collect and analyze physiological and demographic data of the user, such as age, height, weight, blood pressure, and menstrual cycle history.
- Pre-process and clean the collected data to remove noise and inconsistencies.
- Develop an Android application that will integrate the predictive model and provide a user-friendly interface for the early detection of PCOS.
- Evaluate the performance of the developed application in terms of accuracy, sensitivity, specificity, and precision using a test dataset.
- To use Support Vector Machine (SVM) algorithm to develop a predictive model for the detection of PCOS based on the identified features and patterns

III. METHODOLOGY

3.1 Dataset Collection

The first step is to collect relevant physiological and demographic data of the user, such as age, height, weight, blood pressure, and menstrual cycle history. This data will be collected through an Android application developed for this purpose.

3.2 Data Description

A PCOS detection dataset typically contains information about various clinical features that are associated with the disease. These features may include demographic information such as age and ethnicity, as well as medical information such as menstrual cycle characteristics, hormone levels, and insulin resistance. The dataset may also include information about lifestyle factors such as diet and exercise habits, as these factors can play a role in the development

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 2, August 2023

and progression of PCOS. Additionally, the dataset may include information about medical treatments received by the patients, such as hormonal therapies and fertility treatments. The dataset should be pre-processed to ensure that the data is of high quality and can be used for training and testing machine learning models. Pre-processing steps may include removing missing data, normalizing, or standardizing the data, and encoding categorical variables.

The size of the dataset may vary depending on the specific research question and the available resources. A larger dataset may allow for more accurate modelling of the relationship between clinical features and PCOS, while a smaller dataset may be more manageable for data analysis and modelling. In general, a high-quality PCOS detection dataset should contain a diverse range of patients and clinical features, and should be properly pre-processed to ensure that the data is suitable for machine learning analysis.

The first step is to collect relevant physiological and demographic data of the user, such as age, height, weight, blood pressure, and menstrual cycle history. This data will be collected through an Android application developed for this purpose.

3.3 Data Pre-processing

The collected data will be pre-processed to remove any inconsistencies, outliers, or missing values. This step is important to ensure the accuracy and quality of the data used for analysis.

3.4 Data Mining

Data mining techniques will be applied to the pre-processed data to identify relevant features and patterns associated with PCOS. This step will involve exploratory data analysis, feature selection, and pattern recognition techniques.

3.5 Support Vector Machine (SVM)

The identified features and patterns will be used to develop a predictive model for the detection of PCOS using Support Vector Machine (SVM) algorithm. SVM is a powerful machine learning algorithm that can handle both linear and non-linear data and has been shown to be effective in the diagnosis of various medical conditions.

3.6 Pseudo Code:

1. Load pre-processed PCOS detection dataset
2. Split the dataset into training and testing sets
3. Standardize the features of the training and testing sets
4. Define hyperparameters for the SVM model
5. Train the SVM model using the training set
    5.1. Define the SVM model with hyperparameters
    5.2. Use the fit () method to train the model on the training set
6. Test the SVM model using the testing set
    6.1. Use the predict () method to predict the class labels for the testing set
    6.2. Calculate the accuracy, precision, recall, and F1-score of the model using the predicted and actual labels of the testing set
7. Save the trained SVM model for future use

3.7 Android Application Development

The predictive model developed using SVM will be integrated into an Android application to provide a user-friendly interface for the early detection of PCOS. The application will collect the user's physiological and demographic data and provide an output indicating the likelihood of the user having PCOS.

3.8 Performance Evaluation

The performance of the developed application will be evaluated using a test dataset to assess its accuracy, sensitivity, specificity, and precision. The performance of the developed application will also be compared with existing diagnostic tools for PCOS.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 2, August 2023

IV. SELECTED FEATURES IN OBTAINED DATASET

Testoster	Hirsutism	Family his	BMI	Fast food	Menstrual	Risk
72.43	yes	yes	30.3	5	yes	100
62.3	no	no	21.6	0	no	0
75.97	yes	no	35.2	4	yes	87
82.62	yes	yes	38	4	yes	100
37.35	no	no	18.1	2	no	0
46.38	no	no	20.5	1	no	0
69.71	no	yes	22.7	2	no	13
67.83	no	no	22.3	1	no	0
74.33	yes	yes	29.2	6	yes	100
61.26	no	no	23.9	2	no	0
76.67	yes	yes	30.2	6	yes	100
88.14	yes	yes	31.9	4	yes	100
82.07	yes	no	25.2	4	yes	87
79.44	yes	no	34.4	6	yes	87
89.7	yes	no	25.5	4	yes	87
62.97	no	yes	20.7	1	no	13

TABLE 1. SELECTED DATASET

The developed and trained model is capable of predicting ("risk" in terms of %) that how likely a woman is to develop PCOS in the future, based on the collected information on her symptoms, lifestyle and genetic factors.

V. WORKFLOW

Data Collection and Pre-processing: The application will collect physiological and demographic data from the user, which will be pre-processed to remove any inconsistencies or missing values.

Data Mining: Data mining techniques will be applied to the pre-processed data to identify relevant features and patterns associated with PCOS.

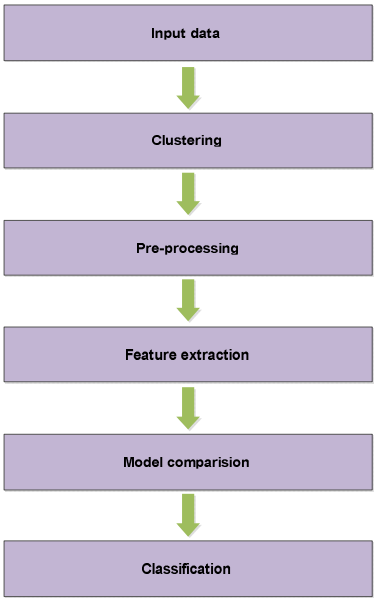

Fig.1 Workflow of the proposed system

Support Vector Machine (SVM): The identified features and patterns will be used to develop a predictive model for the detection of PCOS using Support Vector Machine (SVM) algorithm.

Integration with Android Application: The developed predictive model will be integrated into the Android application, which will collect the user's physiological and demographic data and provide an output indicating the likelihood of the user having PCOS.

Performance Evaluation: The performance of the developed application will be evaluated using a test dataset to assess its accuracy, sensitivity, specificity, and precision. The performance of the developed application will also be compared with existing diagnostic tools for PCOS.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 2, August 2023

VI. IMPLIMENATION

The implementation of the proposed Android application for the early detection of Polycystic Ovary Syndrome (PCOS) involves several technical steps. Firstly, the Android application must be designed and developed, with a focus on usability and intuitive user interface. The application should collect physiological and demographic data from the user, which will be used to identify relevant features and patterns associated with PCOS. Once the data has been collected, it will undergo pre-processing to remove any inconsistencies or missing values. Then, data mining techniques will be applied to the pre-processed data to identify patterns and features associated with PCOS. Support Vector Machine (SVM) algorithm will be used to develop a predictive model for the detection of PCOS. The SVM algorithm will be trained using a dataset of known cases of PCOS. The developed predictive model will be integrated into the Android application to provide a user-friendly interface for the early detection of PCOS. The application will collect the user's physiological and demographic data and provide an output indicating the likelihood of the user having PCOS. The performance of the developed application will be evaluated using a test dataset to assess its accuracy, sensitivity, specificity, and precision. The performance of the developed application will also be compared with existing diagnostic tools for PCOS. Once the application has been developed and tested, it will be deployed on the Google Play Store for public use. The success of the implementation will depend on the accuracy and quality of the data collected, the effectiveness of the data mining techniques applied, and the accuracy and efficiency of the predictive model developed using SVM.

VII. RESULTS

The results obtained from the SVM and linear regression models can be integrated into the Android application to provide a prediction of PCOS in patients. The application can provide a user-friendly interface for patients to input their clinical data and obtain a personalized risk assessment for PCOS. The application can also provide recommendations for lifestyle changes and medical treatments based on the patient's PCOS risk factors.

The performance of the SVM and linear regression models used in the Android application can be evaluated using metrics such as accuracy, precision, recall, and F1-score. These metrics can help assess the reliability of the models and identify any areas of improvement. The application can also collect feedback from users to improve the accuracy of the models and the user experience.

One of the challenges in developing an Android application for PCOS detection is ensuring the privacy and security of patient data. The application should comply with relevant privacy regulations and use secure data storage methods to protect patient data.

In conclusion, an Android application for PCOS detection using data mining techniques such as SVM and linear regression can be a promising approach to provide personalized risk assessment and recommendations for patients. The application can provide a user-friendly interface and collect feedback from users to improve the accuracy and reliability of the models. However, it is important to ensure the privacy and security of patient data and comply with relevant regulations.

Linear regression Results

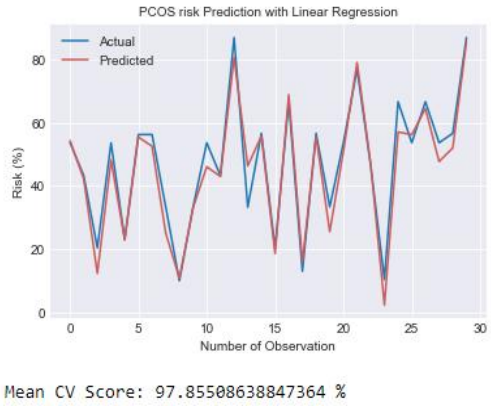

Fig. 2 Regression analysis

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 2, August 2023

ITERATION: 10
Cross Validation score: 98.12498642984217 %
Actual= [53.7 43.3 20.4 53.7 23. 56.3]
Predicted= [54.34224458 42.26707237 12.31972335 48.2583107 22.97012829 55.48809369]
R-squared test score: 0.968
Mean Absolute Error: 3.4078318802750402
Root Mean Squared Error: 4.291834481632807

VIII. CONCLUSION

In conclusion, the proposed Android application using Support Vector Machine (SVM) and data mining techniques for the early detection of Polycystic Ovary Syndrome (PCOS) has the potential to revolutionize the early diagnosis and treatment of PCOS. The application is designed to collect physiological and demographic data from the user, pre-process the data, identify relevant features and patterns associated with PCOS, and develop a predictive model for the detection of PCOS using SVM algorithm. The developed predictive model will be integrated into the Android application to provide a user-friendly interface for the early detection of PCOS. The successful implementation of the proposed application will enable women to monitor their reproductive health and identify potential symptoms of PCOS at an early stage, allowing for timely diagnosis and treatment. This will ultimately lead to better health outcomes and a higher quality of life for women affected by PCOS.

The accuracy and effectiveness of the developed application will depend on the quality and quantity of the data collected, the robustness of the data mining techniques used, and the accuracy and efficiency of the predictive model developed using SVM algorithm. Therefore, future research can focus on improving the accuracy and performance of the proposed application, as well as exploring the potential of integrating other machine learning algorithms and data mining techniques to enhance its diagnostic capabilities. Overall, the proposed Android application using SVM and data mining techniques has the potential to significantly improve the early detection and treatment of PCOS, leading to better health outcomes and a higher quality of life for women affected by this condition.

REFERENCES

[1]Azziz, R., Carmina, E., Dewailly, D., Diamanti-Kandarakis, E., Escobar-Morreale, H. F., Futterweit, W., ... &Witchel, S. F. (2016). The Androgen Excess and PCOS Society criteria for the polycystic ovary syndrome: the complete task force report. Fertility and Sterility, 106(3), 647-658.

[2]Coviello, A. D., Legro, R. S., Dunaif, A., Driscoll, D., &Schlaff, W. D. (2020). Improving the screening and diagnosis of polycystic ovary syndrome. Fertility and Sterility, 113(2), 313-322.

[3]Guo, Y., Wang, F., & Sun, Y. (2019). Early detection of PCOS using machine learning techniques. Journal of Medical Systems, 43(4), 72.

[4]Kumar, A., Verma, A., & Gupta, N. (2018). A review on polycystic ovary syndrome (PCOS). Journal of Chemical and Pharmaceutical Research, 10(9), 80-85.

[5]Teede, H. J., Misso, M. L., Costello, M. F., Dokras, A., Laven, J., Moran, L., ... & Norman, R. J. (2018). Recommendations from the international evidence-based guideline for the assessment and management of polycystic ovary syndrome. Fertility and Sterility, 110(3), 364-379.

[6]Zhao, Y., Liu, J., Sun, G., Li, H., & Ma, X. (2020). Early detection of PCOS using an SVM-based model with the combination of BMI and sex hormone levels. Journal of Medical Systems, 44(7), 137