WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

A Deep Learning Approach for Predicting Polycystic Ovary Syndrome

Mansi Kamble[1], Hemakshi Mandekar[2], Vismaya Nair[3], Janhavi Sangoi[4]

[1, 2, 3]Final Year Student, Computer Enginnering Department, VIVA Institute of Technology, India
[4]Assistant Professor, Computer Enginneering Department, VIVA Institute of Technology, India

A B S T R A C T

Polycystic Ovary Syndrome (PCOS) is a common endocrine disorder in women of reproductive age, affecting approximately 5-10% of the female population. It is a complex and multifactorial disorder with symptoms such as menstrual irregularities, infertility, hirsutism, acne, and obesity. Due to its heterogeneous nature, the diagnosis of PCOS is often challenging and requires a combination of clinical, biochemical, and ultrasound criteria. In recent years, machine learning techniques have been employed for the early detection and diagnosis of PCOS. The time and cost involved in numerous clinical tests and ovary scanning has become a burden to the patients. To address this problem a deep learning-based PCOS prediction system is developed that can predict PCOS in women based on their medical history and symptoms. Deep learning is a subset of machine learning, which is prominently used because of the feature extraction. A Multi-Layer Perceptron (MLP) neural network model is trained for classification, which takes the pre-processed data as input and outputs the probability of having PCOS. The proposed approach of PCOS prediction using deep learning technique is a promising solution for the early detection and diagnosis of PCOS. The implemented model can be integrated into clinical decision support systems to assist healthcare professionals in making accurate and timely diagnosis.

Keywords: Accuracy, Deep Learning Model, Detection, Parameters, PCOS.

1. Introduction

There are many kinds of disorders suffered by people and to predict them various systems are available. Polycystic ovary syndrome (PCOS) is also one kind of disorder suffered by women. It is a hormonal endocrine disorder found among women of reproductive age. Over five million women worldwide are suffering from Polycystic ovary syndrome (PCOS) [3]. Polycystic ovary syndrome (PCOS) is a condition in which the ovaries produce an abnormal amount of androgens, male hormones that are usually present in women in small amounts. The name polycystic ovary syndrome describes the numerous small cysts (fluid-filled sacs) that form in the ovaries. However, some women with this disorder do not have cysts, while some women without the disorder do develop cysts.

In some cases, a woman doesn't make enough of the hormones needed to ovulate. When ovulation doesn't happen, the ovaries can develop many small cysts. These cysts make hormones called androgens. Women with PCOS often have high levels of androgens. This can cause more problems with a woman's menstrual cycle. And it can cause many of the symptoms of PCOS. For women, the reproductive system is one of the most vital organs. This disorder can create extreme difficulties or even failure of ovulating frequently and therefore, makes it difficult for the patient to get pregnant. Approximately 5-10 % of reproductive age (15-49 years) women suffer from this problem. The most common symptoms of this disorder may include missed periods, irregular periods, or very light periods, it affects in a way that ovaries become large or may contain many cysts, it can also cause excess body hair, including the chest, stomach, and hirsutism, can cause weight gain, especially around the abdomen, Acne or oily skin [2]. There are many systems to predict PCOS but that are based on images for images it requires sonography reports. Also, these can be costly and probabilities of having PCOS can be less in some cases.

Deep Learning is a highly dynamic technology that has the potential to address various challenges across different fields. PCOS is a multifactorial and polygenic condition. Deep learning is capable of "learning" features from very large amounts through clinical practice to diagnose this disorder. In the healthcare sector, deep learning is assisting practitioners and researchers in uncovering hidden insights from vast amounts of data, enabling the medical industry to operate more efficiently [7]. It is also facilitating doctors in accurately diagnosing diverse medical conditions and providing optimal treatment plans, leading to better healthcare outcomes.

2. Related Work

This section describes the various methodologies used in prior studies to identify and classify Polycystic ovary syndrome (PCOS). Malik Mubasher Hassan, Tabasum Mirza [1], In this research paper machine learning approaches like Support Vector Machine, CART, Naive Bayes Classification, Random Forest and Logistic Regression to predict PCOS based on the patient's clinical data are used by authors. The validation metrics indicate the highest i.e. Random Forest algorithm in the diagnosis of PCOS on giving data. Priyanka R. Lele , Anuradha D. Thaka [2], This paper presents an approach where classification of PCOS will use physical symptoms and sonograms. The results of only physical symptoms are presented here. Amongst all the

* Corresponding author. Tel.: +0-000-000-0000 ; fax: +0-000-000-0000.
E-mail address: author@institute.xxx

algorithms K-star algorithm is out performing in all the performance measures. Dr. Ashok Munjal, Dr. Rekha Khandia, Brijraj Gautam [3], This paper encompasses the use of genetic algorithms and a Machine learning approach for selection of major attributes (the signs and symptoms). Extra Tree classifier gives better accuracy of 88% as compared to Decision Tree and Random Forest classifier. Pushkarini H, M A Anusuya [4], In this paper, three trained models are evaluated using performance measures and the results are compared to find the best model for PCOS risk prediction. Linear regression, KNN and Random forest models are implemented and evaluated on various performance metrics. Mr. Abdul K, Shreasta M, Vani N, Vathsala K, Vidhya Shree [5], used machine learning algorithms such as AdaBoost, Gradient Boosting, KNN, Random Forest, and Logistic Regression to diagnose PCOS based on patient clinical data. The data was analyzed, and the algorithms' accuracy and precision were validated. The Random Forest algorithm in the diagnosis of PCOS on given data has the maximum accuracy.Vikas B, B. S. Anuhya, Manaswini Chilla, Sipra Sarangi [6], In this paper, an attempt has been made to compare the accuracies and other performance measures of the prior mentioned data mining techniques such as Naive Bayes, Decision Tree and Artificial Neural Networks are applied to predict whether a person is likely to have PCOS or not through. The Naïve Bayes algorithm provides with optimum accuracy followed by ANN backpropagation algorithm. Subrato Bharati, Prajoy Podder, M. Mondal [7], The paper focuses on the prediction of PCOS by selecting the most important attributes of PCOS patients for the given dataset by using feature selection method. Among these, RFLR has the best testing accuracy when 40-fold cross validation is used to split the data into testing and training portions. Amsy Denny, Anita Raj, Ashi Ashok, Maneesh Ram C, Remya George [8], This paper proposes a system for the early detection and prediction of PCOS from optimal and minimal but promising clinical and metabolic parameters. Classification of PCOS with the feature set transformed with Principal Component Analysis (PCA) is done using various machine learning techniques such as Naïve Bayes classifier method, logistic regression, K-Nearest neighbor (KNN), Classification and Regression Trees (CART), Random Forest Classifier, Support Vector Machine (SVM) in Spyder Python IDE. Results revealed that the most suitable and accurate method for the PCOS prediction is RFC with an accuracy of 89.02%. Kinjal Raut, Chaitrali Katkar, Suhasini A. Itkar [9], This paper focuses on prediction of PCOS, this includes first selection of most appropriate attributes using the feature selection method from the dataset. The algorithms used to construct the model are SVC, Decision Tree, Logistic Regression, K Nearest Neighbor, Random Forest, XGBRF and CatBoost Classifier. These algorithms are then compared to check the accuracy. Muhammad Inan, Rubaiath Ulfath, Fahim Alam, Fateha Bappee, Rizwan Hasan [10], In this paper, the data has been resampled by combination of both SMOTE (Synthetic Minority Oversampling Techniques) & ENN (Edited Nearest Neighbour) to solve the class imbalance problem in medical domain datasets and data outliers' issues. Different classifiers such as KNN, SVM, RF, Adaboost, NB and MLP individually to generate recall, precision and compared each of these results with the XGBoost model. Palak Mehrotra,Jyotirmoy Chatterjee,Chandan Chakraborty [11], In this paper the data is gather from the study conducted on the patients coming to Ghosh Dastidar Institute for Fertility Research (GDIFR), Kolkata between March 2010 and April 2011. To classify the selected attributes two algorithms Bayesian and Logistic Regression (LR) classifiers are used. The two classifiers used provided consistent results in all folds. Based on four clinical features: FSH, LH, BMI, Cycle Length the occurrence probability for PCOS was predicted. Vaidehi Thakre, Shreyas Vedpathak, Kalpana Thakre and Shilpa Sonawani [12], This paper proposes a system which can help in early detection and prediction of PCOS treatment from an optimal and minimal set of parameters. To detect whether a woman is suffering from PCOS, 5 different machine learning classifiers like Random Forest, SVM, Logistic Regression, Gaussian Naive Bayes, K Neighbours have been used.

3. Research Objectives

The current approach for predicting Polycystic Ovary Syndrome (PCOS) involves using an excessive number of parameters. To improve accuracy, it would be beneficial to use feature selection techniques and only consider the most important features. There are also some systems which can predict PCOS but that are based on images for which it requires sonography reports. Also, these can be costly and probabilities of having PCOS can be less in some cases. So, this implemented system aims to solve these problems and can be used for early prediction of PCOS based on different features and also it can be cost efficient. There is no system available which can predict PCOS using deep learning based on features. The implemented system will be able to predict PCOS accurately as there are minimal features which provide more accuracy.

4. Proposed System & Methodology

4.1 Data Collection:

The process of developing a deep learning system begins with the processing of the dataset. "Polycystic ovary syndrome (PCOS)"[16], a dataset from website, which mainly consists of 541 data among which 177 women are suffering from PCOS and 364 are normal one who are not suffering from PCOS. There are 43 different attributes in dataset like period cycle, BMI, hair loss, etc. Upon successful data collection feature selection technique has been carried out in which 19 parameters are selected from 43 attributes. The 19 parameters which selected are Age, BMI, Cycle, Cycle length, Marriage, Pregnant, number of abortions, Waist: Hip Ratio, TSH, PRL, Vit D3, RBS, Hair Growth, Skin Darkening, Hair Loss, Pimples, Fast food, Exercise.

4.2 System Block Diagram:

The Fig.1 depicts the block diagram of the proposed system. First, the dataset is collected from Kaggle, then data pre-processing is carried out, which consist of data cleaning and labelling. Further the feature selection technique has been carried out in which 19 parameters are selected from 43 attributes. The prediction of PCOS will be based on these 19 parameters. Afterward, the dataset is split into an 80:20 ratio, where 80% of the data is used for training and the remaining 20% is used for testing the model. Classification is performed by applying a deep learning algorithm, specifically Multilayer Perceptron (MLP). The model is tested, and prediction of Polycystic Ovary Syndrome (PCOS) is carried out to determine whether the user has PCOS or not.

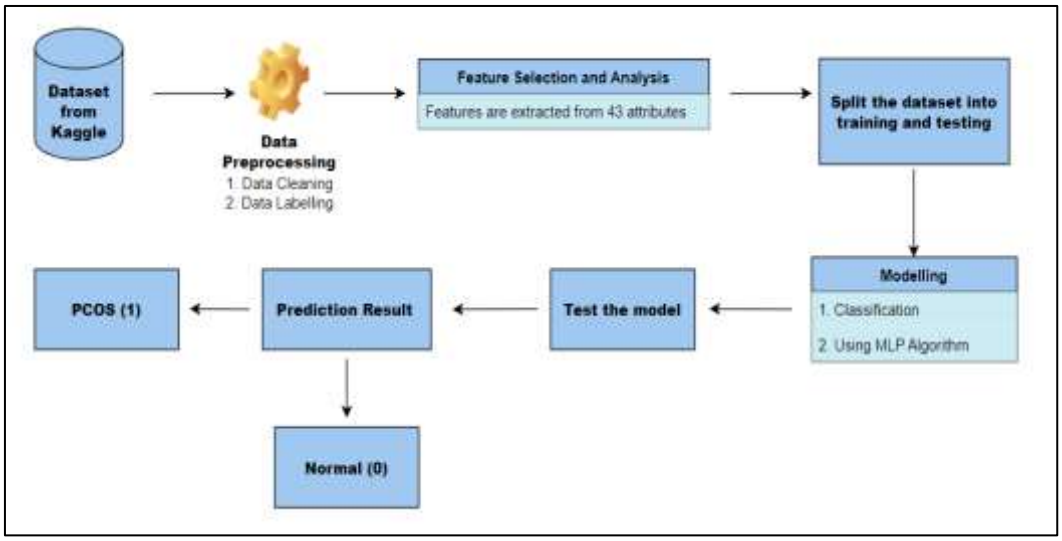

Fig. 1 Block Diagram for PCOS Prediction

4.3 Multilayer Perceptron (MLP) Architecture:

The architecture of the MLP system is shown in Fig. 2. A multilayer perceptron (MLP) is a type of artificial neural network (ANN) consisting of multiple layers of nodes, or neurons, that are interconnected in a feedforward manner. It consists of three types of layers: an input layer, one or more hidden layers, and an output layer. The input layer is the first layer of the MLP, and its neurons receive the input data. Each neuron in the input layer corresponds to one feature of the input data. The hidden layers are located between the input layer and the output layer. Each hidden layer contains multiple neurons that receive inputs from the neurons in the previous layer and apply a nonlinear activation function to produce an output. The output layer is the final layer of the MLP, and its neurons produce the output of the network. The number of neurons in the output layer depends on the type of task the MLP is designed for. An MLP classifier is defined with a single hidden layer of 90 neurons using the MLPClassifier () function. The activation function used for the hidden layer is 'relu', and the solver used is 'lbfgs'. The MLP classifier is then trained on the training set using the fit () function. A new data point is then created for prediction using the trained classifier. This data point is in the form of a 1x19 array, with values for the 19 features. The predict () function is then used to predict the class of this data point, which is either 0 or 1.

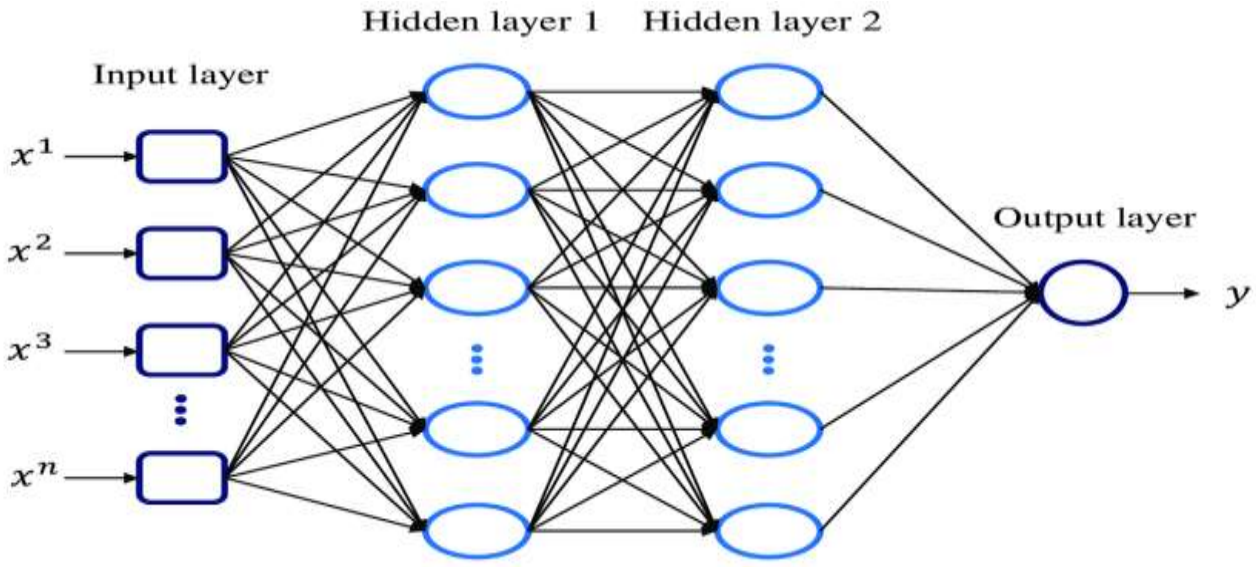

Fig. 2 MLP Architecture

5. Results & Analysis

The system takes input from the user through a form that asks for specific parameters related to the symptoms of PCOS. To showcase the practical application of the PCOS prediction system, a web page has been developed. The web page includes a home page, a test page that contains a form for user input and displays the prediction results, and a remedy page that provides useful remedies to overcome PCOS. Separate pages have been created for both positive and negative prediction results.

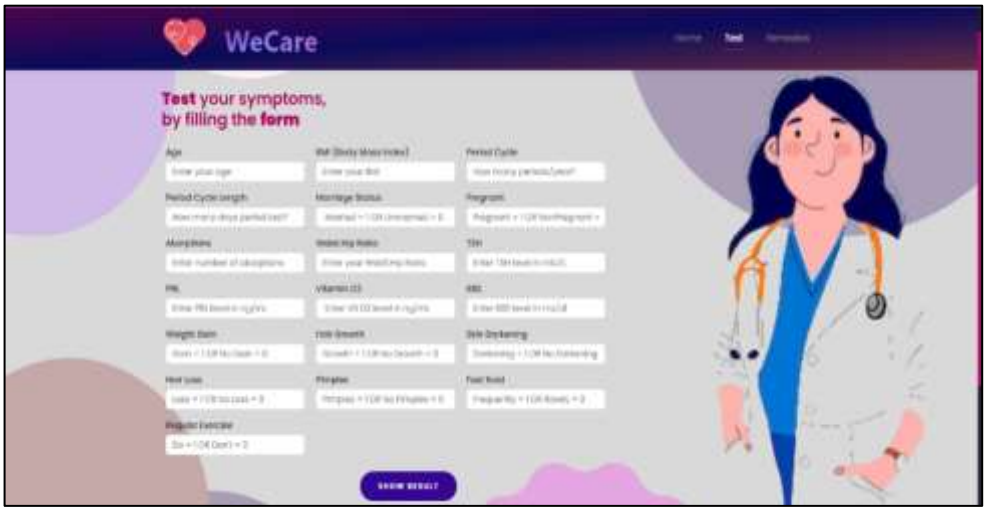

Fig. 3 Test Page

In Figure 3, the user needs to fill the form which contains 19 parameters to predict PCOS.

Fig. 4 Test Page with filled parameters

In Figure 4, the user fills the form containing the 19 parameters to predict PCOS.

Fig. 5 Result page with PCOS

In Figure 5, if the result is pcos, then it will display the above page which shows that the user has pcos along with a diet chart to refer.

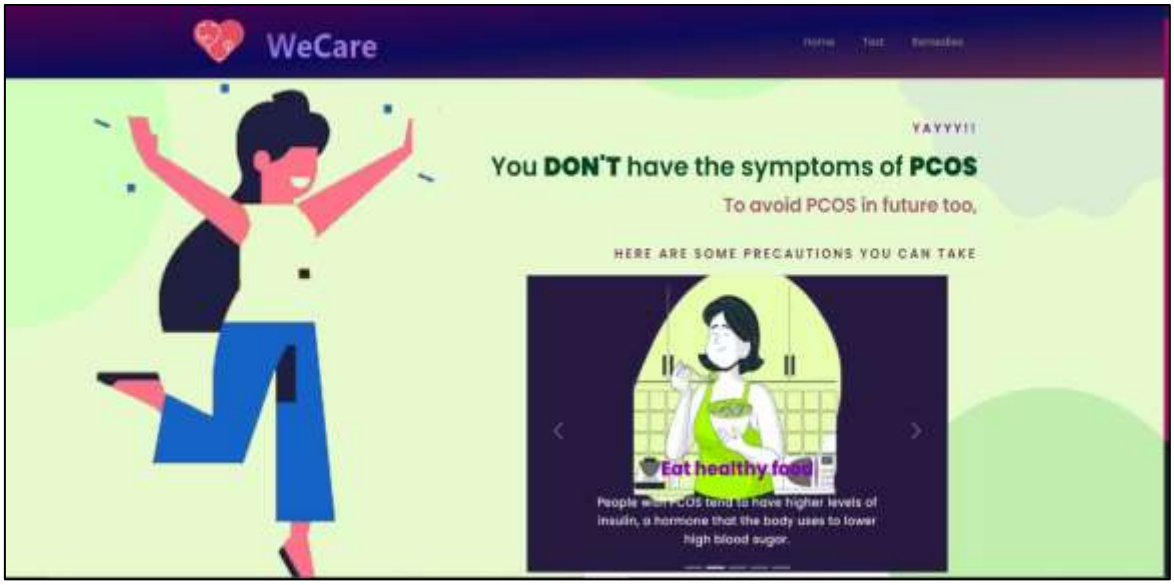

Fig. 6 Result page without PCOS

In Figure 6, if the result is no pcos, then it will display the above page along with the precautions to be taken in future.

After training the model, a classification report is generated using sklearn which contains precision, recall, f1-score, and support as shown in the table 3.1. A classification report provides a comprehensive evaluation of the performance of a classification model. it helps to identify the strengths and weaknesses of the model, and provides a better understanding of how well the model is performing for each class. This information can be used to fine-tune the model or to compare different models and select the best one for the task at hand.

Table 1 – Accuracy Table

	Precision	**Recall**	**F1-score**	**Support**
0	0.97	0.96	0.96	100
1	0.96	0.97	0.97	100
accuracy			0.96	200
macro avg	0.97	0.96	0.96	200
weighted avg	0.97	0.96	0.96	200

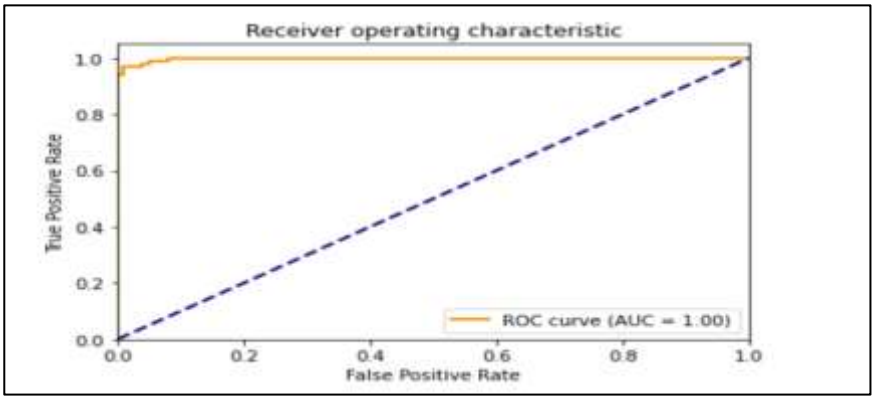

Fig. 7 ROC Graph

Fig 7, shows the graph of ROC curve which represents the trade-off between the true positive rate (TPR) and false positive rate (FPR) of a binary classifier.

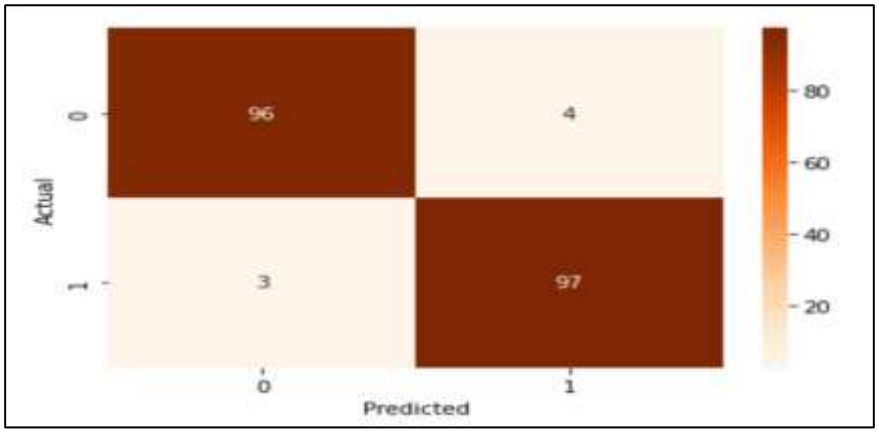

Fig. 8 Confusion Matrix of MLP Model

Fig 8, represents a confusion matrix where the rows depict the actual classes and the columns represent the predicted classes.

Table 2 shows the accuracy measures of MLP model. One metric for assessing classification models is accuracy. Accuracy is the proportion of correct predictions made by the model. Accuracy for MLP model is evaluated as follows:

$$Accuracy = \frac{No.\ of\ correct\ predictions}{Total\ no.\ of\ predictions}$$

A common statistic for evaluating the accuracy of an object detection model is the Mean Average Precision (mAP). The mAP value is estimated by calculating the Average Precision (AP) for each class and then averaging over a number of classes.

The mAP of MLP model is evaluated as follows:

$$mAP = \frac{1}{N}\sum_{i=1}^{n} AP$$

The MLP model gives an accuracy of 0.96.

The F1-score is an error metric that calculates the harmonic mean of precision and recall to evaluate model performance. It offers reliable results for both balanced and unbalanced datasets and considers the model's precision and recall capabilities. F1-score of a model is evaluated as follows:

$$F1\ Score = 2 * \frac{precision * recall}{precision + recall}$$

For object detection model F1-score can also be evaluated as follows:

$$F1\ Score = 2 * \frac{mAP * mAR}{mAP + mAR}$$

where, Mean Average Recall (mAR) is the mean of Average Recall (AR) values.

MLP model gives F1-score of 0.96.

Table 2 – Model Accurateness

Model	Accuracy	F1-score
MLP	0.96	0.96

6. Conclusion

The research is focused on Polycystic Ovary Syndrome (PCOS) is a condition which causes hormonal disorder in women in their reproductive age. It is extremely difficult to diagnose PCOS due to the heterogeneity of symptoms. The time and cost involved in numerous clinical tests and ovary scanning has become a burden to the patients. Moreover, there are various systems which are based on machine learning models that require more than 23 features to predict whether women are suffering from PCOS or not. These features contain some kind of sonography reports so this can again be a time-consuming job. Thus, the implemented system can predict Polycystic Ovary Syndrome (PCOS) based on 19 features. If the woman is suffering from PCOS it will provide output as positive PCOS (1) and negative if Normal (0). In this system Multilayer Perceptron algorithm has been used which is a deep learning model. In this algorithm there is a single hidden layer with 90 neurons so that the model will be able to learn different patterns and can predict accurately based on parameters. By using the Multilayer perceptron algorithm, the accuracy achieved is 96%. The use of this kind of model will help the doctors for the early screening of the patients who are more likely to develop the disease. The implemented system can also show different remedies which will include healthy diets and different exercises to overcome PCOS.

References

[1] M. Hassan, T. Mirza, "Comparative Analysis of Machine Learning Algorithms in Diagnosis of Polycystic Ovarian Syndrome", International Journal of Computer Applications, Vol. 175 No.17, 2020.

[2] P. Lele, A. Thakare, "Comparative Analysis of Classifiers for Polycystic Ovary Syndrome Detection using Various Statistical Measures", IJRET, Vol. 9 Issue 03, 2020.

[3] Dr. A. Munjal, Dr. R. Khandia, B. Gautam, "A Machine Learning approach for selection of Polycystic Ovarian Syndrome (PCOS) attributes and comparing different classifier performance with the help of Weka and Pycaret.", International Journal ofF Scientific Research, Vol. 9 Issue - 12, 2020.

[4] H. Pushkarini, M. Anusuya, "A Prediction Model for Evaluating the Risk of Developing PCOS", IJRET, Vol 07 Issue 09, 2020.

[5] Khadar, M. Shreasta2, N. Vani, K. Vathsala, V. Shree, "Diagnosis of Polycystic Ovary Syndrome using Machine Learning Algorithms", IJARSCT, Vol 2 Issue 06, 2022.

[6] Vikas, B. Anuhya, M. Chilla, S. Sarangi, "A Critical Study of Polycystic Ovarian Syndrome (PCOS) Classification Techniques", IJCEM, Vol. 21 Issue 4, 2019.

[7] S. Bharati, P. Podder, M. Mondal, "Diagnosis of PCOS Using Machine Learning Algorithms", IEEE, Vol. 2 Issue 04, 2020.

[8] A.Denny, A. Raj, A. Ashok, M. Ram, R. George, "i-HOPE: Detection And Prediction System For Polycystic Ovary Syndrome (PCOS) Using Machine Learning Techniques", IEEE, 2021.

[9] K. Raut, C. Katkar, S. Itkar, "PCOS Detect using Machine Learning Algorithms", IRJET, Vol. 09 Issue 01, 2022.

[10] M. Inan, R. Ulfath, F. Alam, F. Bappee, R. Hasan, "Improved Sampling and Feature Selection to Support Extreme Gradient Boosting for PCOS Diagnosis.", IEEE, 2021.

[11] P. Mehrotra,J. Chatterjee, C. Chakraborty, "Automated Screening of Polycystic Ovary Syndrome using Machine Learning Techniques." IEEE, Vol. 03 Issue 05, 2020.

[12] V. Thakre, S. Vedpathak, K. Thakre and S. Sonawani, "PCOcare: PCOS Detection and Prediction using Machine Learning Algorithms.", IJRET, Vol. 12 Issue 01, 2021.

[13] https://www.hopkinsmedicine.org/health/polycystic-ovary-syndrome-pcos last accessed on: 2/10/2022.

[14] https://en.wikipedia.org/wiki/Polycystic_ovary_syndrome. last accessed on: 27/9/2022.

[15] https://www.mayoclinic.org/diseases-conditions/pcos/symptoms-causes last accessed on: 1/10/2022.

[16] https://www.kaggle.com/datasets/prasoonkottarathil/polycystic-ovary-syndrome-pcos last accessed on: 25/09/2022.