JOURNAL OF SOFT COMPUTING AND DATA MINING
e-ISSN: 2716-621X

Vol. 5 No. 2 (2024) 72-82
https://publisher.uthm.edu.my/ojs/index.php/jscdm

JSCDM

Data-driven Clinical Decision Support System Using Neural Network Topology Optimization for PCOS Diagnosis

Leema Nelson[1]*

[1] *Chitkara University Research & Innovation Network (CURIN),*
Chitkara University Institute of Engineering & Technology, Chitkara University, Punjab, INDIA

*Corresponding Author: leema.nelson@gmail.com
DOI: https://doi.org/10.30880/jscdm.2024.05.02.006

Article Info

Received: 19 May 2024
Accepted: 9 December 2024
Available online: 18 December 2024

Keywords

Neural network, particle swarm optimization, polycystic ovary syndrome, correlation-based feature selection, diagnosis

Abstract

Polycystic ovarian syndrome (PCOS) is a prevalent endocrine disorder that affects women of reproductive age worldwide. Chronic oligo- or anovulation, clinical hyperandrogenism, and polycystic ovaries characterize it. Long-term complications include endometrial cancer, infertility, obesity, and type 2 diabetes mellitus. Early diagnosis and treatment are crucial to reduce the incidence of these complications. This work aims to develop a Clinical Decision Support System (CDSS) utilizing Multi-Objective Particle Swarm Optimization (MOPSO) - backpropagation (BP) with artificial neural networks (ANN) to determine the presence of PCOS. CDSS comprises three subsystems: preprocessing, training, and classification. The preprocessing subsystem manages the missing values and performs correlation-based feature selection with a threshold of 0.25. The training subsystem employs an ANN trained with BP and MOPSO. The local best values are obtained from the BP training seed MOPSO, which has two objective functions: minimizing the mean square error and achieving faster convergence without stagnation. The PCOS dataset from the Kaggle repository is used in the experiments. The developed CDSS achieved an overall accuracy of 92.02%, with 87.27% sensitivity and 94.44% specificity. The CDSS is a valuable second-opinion tool for junior gynecologists in diagnosing PCOS, offering a robust and accurate method for early detection and intervention.

1. Introduction

Polycystic ovarian syndrome is a heterogeneous hormonal condition that affects women of reproductive age and is associated with a range of health problems, such as hypertension, dyslipidemia, insulin resistance, hyperandrogenemia, and type 2 diabetes mellitus [1]. According to reports, globally, PCOS accounts for 5.5% and 12.6% of women in the age range of 17 to 45 years [2]. The prevalence of PCOS is estimated between 8.2% - 22.5% in India, depending on the diagnostic criteria used [3, 4]. PCOS causes infertility, obesity, lifestyle changes, and a family history of PCOS and endometrial cancer. PCOS is a lifestyle disorder related to modernization that fails to follow traditional lifestyles. Different symptoms can be used to diagnose this hormonal disorder, such as missed or irregular periods, hair loss, acne, depression, hirsutism, weight gain, and growth of ovarian cysts. PCOS impacts health and quality of life for women. Early diagnosis and treatment are commonly accomplished to reduce long-term effects.

NIH Criteria [6]: Diagnosis of PCOS based on oligoovulation (irregular menstrual cycles) and excess androgen levels. This criterion originated from a PCOS meeting at the National Institutes of Health (NIH) in 1990, where oligo-ovulation and excess androgen were identified as key diagnostic markers. Rotterdam Criteria [7]:

Established in 2003 by the European Society of Human Reproduction and Embryology and the American Society for Reproductive Medicine. The presence of two out of three factors, oligo/anovulation, polycystic ovaries, and androgen excess, is used for diagnosing PCOS. This criterion offers a broader scope by considering additional factors, such as polycystic ovaries, oligo/anovulation, and androgen excess. AE-PCOS Criteria [1]: Published in 2006 by the Androgen Excess Society and PCOS Society. The diagnosis of PCOS is based on androgen excess and the presence of one of the two symptoms. This provides an alternative approach by emphasizing androgen excess alongside specific symptoms for diagnosing PCOS. Updated Rotterdam Criteria (2018) [8]: International evidence-based guidelines for PCOS updated the Rotterdam criteria in 2018 with minor modifications. Retains the essence of the original Rotterdam criteria, but may include refinements or adjustments based on updated research and clinical insights. Each of these criteria offers a distinct perspective on diagnosing PCOS, considering various combinations of symptoms and markers, such as oligo/anovulation, androgen levels, polycystic ovaries, and specific associated symptoms

Machine learning (ML) and artificial intelligence (AI) have become effective tools for resolving complicated issues in various fields. ML plays a major role in clinical decision-making and diagnosis [9]. The fundamental challenge for the global healthcare system is to provide accurate and accessible diagnosis. ML allows us to build a classifier model associated with a set of features for a disease [27]. The combination of data science and machine learning provides an accurate diagnostic model. This combination analyzes clinical data to determine the relationship between diagnostic features, design a predictive model and test the model using new samples [26].

This research presents a method for improving the accuracy of a Clinical Decision Support System (CDSS) for diagnosing PCOS. The objective of this study is to address the stagnation issue in PSO and enhance CDSS accuracy by introducing a technique known as MOPSO-BP. This method combines Multi-Objective Particle Swarm Optimization (MOPSO) with backpropagation (BP) to train an Artificial Neural Network (ANN) for PCOS diagnosis. The study uses BP as the starting point for MOPSO during ANN training to find the local best values and perform a global search to determine the optimal set of ANN weights, thereby improving the overall performance of the CDSS.

This paper's main structure is organized as follows: In Section 2 of this article, a review of earlier studies is provided. The materials and procedures used in this study are described in Section 3. The framework of the proposed system is presented in section 4. Section 5 discusses the experimental results. Finally, Section 6 concludes this paper and offers a framework for further study.

2. Literature Review

Several studies have reviewed the ANN literature, which has been used to develop classification models using bio-inspired algorithms for clinical data. Rahman et al. [10] have developed a web-based machine-learning method for predicting PCOS. This study analyzed a dataset containing 541 patient records obtained from the Kaggle repository to identify patterns in a particular disorder using machine learning models. This study evaluated the accuracy, specificity, sensitivity, and precision of various models, including Logistic Regression, Decision Tree, AdaBoost, Random Forest, and Support Vector Machine. The Mutual Information model was employed for feature selection and comparison, and AdaBoost and Random Forest achieved the highest accuracies of 94%. Logistic Regression is easy to interpret but may struggle with complex nonlinear relationships. Decision Trees are simple to understand and visualize but can overfit with noisy data. AdaBoost combines weak learners for better accuracy but is sensitive to outliers. Random Forest is robust against overfitting and noise; however, careful hyperparameter tuning is required. SVM is appropriate for high-dimensional data and can handle nonlinear relationships; however, it is computationally intensive. AdaBoost and Random Forest achieved an accuracy of 94%, highlighting their effectiveness in predicting PCOS.

Tiwari et al. [11] have developed a smart polycystic ovary syndrome diagnostic system (SPOSDS) using machine learning techniques. This research examines PCOS diagnosis using a clinical dataset provided by Kottarathil, which is available in the Kaggle repository. This study employed non-invasive screening parameters to assess various machine learning techniques for screening patients with PCOS without relying on invasive diagnostics. The results of the experiments demonstrate that the Random Forest (RF) method achieves other prominent machine learning algorithms with an accuracy of 93.25%. Additionally, the out-of-bag (OOB) error was used to gauge the prediction performance of the RF. RF accuracy and OOB error metrics show its potential as a reliable CDSS tool for PCOS diagnosis.

Suha et al. [12] conducted a study investigated the dominant features and data-driven detection of polycystic ovary syndrome using a modified stacking ensemble machine learning technique. This study proposes a modified ensemble machine learning classification approach that uses state-of-the-art stacking techniques and patient symptom data to identify PCOS. The approach employs five traditional ML models as base learners and one bagging or boosting ensemble ML model as the meta-learner of the stacked model. Furthermore, three different types of feature-selection strategies were used to select varying numbers and combinations of attributes. The proposed technique was trained, tested, and assessed using different feature sets, and the results showed that it

J. of Soft Computing and Data Mining Vol. 5 No. 2 (2024) p. 72-82

significantly improved the accuracy compared with existing ML-based techniques for all varieties of feature sets. Among the various models investigated to classify PCOS and non-PCOS patients, the stacking ensemble model with a Gradient Boosting classifier as the meta-learner outperformed the others with 95.7% accuracy when using the top 25 features selected using the Principal Component Analysis feature selection technique.

Jain et al. [13] employed the exploratory data analysis (EDA) technique to evaluate the features of a tabular dataset in a designated pipeline. Subsequently, a range of machine learning algorithms, such as Gaussian Naive Bayes (GNB), k-nearest neighbors (kNN), support vector machine (SVM) with linear and RBF kernels, Random Forest (RF), and dense neural network (DNN) classifiers, are utilized to attain a high accuracy of 97%. Furthermore, the explanation provided by LIME and SHAP enhanced the model, resulting in a reliable, interpretable, and self-explanatory deep learning diagnostic model. By conducting dataset analysis using explainable artificial intelligence (XAI), insulin level was found to be the most significant factor affecting PCOS, followed by follicle no (R) and follicle no (L). Consequently, this research significantly contributes to the identification of the factors responsible for PCOS disorders using XAI.

Their comparative performance demonstrated the superiority of machine learning models over traditional diagnostic methods. Traditional methods frequently involve invasive diagnostics, whereas machine-learning models offer noninvasive, precise, and dependable alternatives [14], [25]. This analysis presents a comprehensive examination of the current leading strategies for automating Polycystic Ovary Syndrome (PCOS) diagnosis. The reviewed studies highlight the use of various machine learning models to enhance diagnostic accuracy and dependability. The potential of a Clinical Decision Support System (CDSS) using these methods demonstrates the possibility of improved clinical outcomes and a more streamlined PCOS diagnosis. These results provide a method for the development of a Multi-Objective Particle Swarm Optimization-based BP Neural Network (MOPSO-BP) model, which builds upon the strengths of existing methodologies while addressing their shortcomings.

3. Materials and Methods

This section describes in detail the methodologies and tools used to create the CDSS. The main objective of this research is to develop a CDSS that can accurately diagnose PCOS by employing a combination of Multi-Objective Particle Swarm Optimization (MOPSO) and backpropagation (BP) with artificial neural networks (ANN). The methodology uses three distinct phases: preprocessing, training, and classification, as shown in Fig 1.

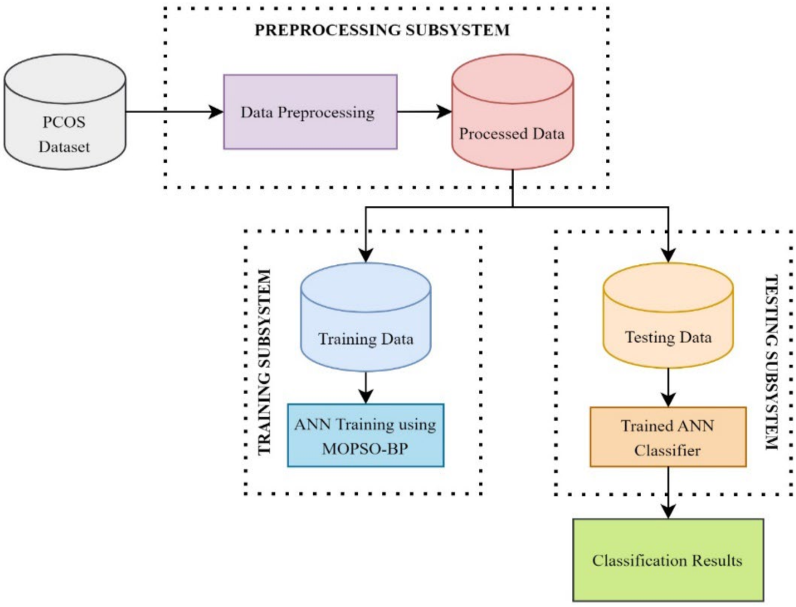

Fig. 1 *System framework of ANN-MOPSO-BP model*

Initially, the preprocessing subsystem is used to address the missing data and perform correlation-based feature selection using a threshold of 0.25. The training subsystem then utilizes BP and MOPSO to train an ANN, with MOPSO optimizing two objective functions: minimizing the mean square error and achieving faster convergence without stagnation. Finally, the classification subsystem provides a trained ANN to accurately diagnose PCOS. MATLAB R2021a is used to carry out the implementations. The supervised machine learning ANN algorithm and metaheuristic- MOPSO method are described in detail.

3.1 Artificial Neural Network (ANN)

In this study, a CDSS is developed using an ANN prediction model. This model uses a multilayer feed-forward network with a single hidden layer. Each layer in the multilayer feed-forward network processes the data received from the layer before it and sends it to the layer after. The input layer contained $\boldsymbol{n}$ nodes, the hidden layer contained $\boldsymbol{H}$ nodes, and the output layer contained $\boldsymbol{O}$ nodes.

Step 1: Calculate the j^{th} hidden node output using equation (1).

$$f(y_j) = {}^{1}\!\Big/\!\left(1 + \exp\left(-(\textstyle\sum_{i=1}^{n} w_{ji}\, x_i - \theta_j)\right)\right), \quad \text{j= 1, 2..., H} \tag{1}$$

where x_i is the i^{th} input, w_{ji} is the weight from the i^{th} input node to the j^{th} hidden node, and θ_j is the hidden layer threshold.

Step 2: Calculate the output layer (z_k) output using equation (2).

$$z_k = \textstyle\sum_{j=1}^{H} w_{kj}\, f(y_j) \qquad k = 1,2, \dots \dots, O \tag{2}$$

The weight from the j^{th} hidden node to the k^{th} output node is denoted by w_{kj}.

Step 3: The number of hidden layer nodes is calculated using equation (3),

$$H = (2n + 1) \tag{3}$$

Step 4: Calculate the Mean Square Error (MSE) E_k of the network using equation (4).

$$E_k = \textstyle\sum_{i=1}^{O} (z_i^k - C_i^k)^2 \tag{4}$$

Step 5: Calculate the error (Δw_k) in the network in response to the change in weights and biases, as shown in equation (5)

$$\Delta w_k = \eta({z^k}_i - {C^k}_i)x_i \tag{5}$$

Step 6: Update the error in the network using the Gradient Descent BP algorithm, which is represented in equation (6)

$$\Delta w_k = -\eta . g_k \tag{6}$$

where η is the learning rate, and g_k is the current gradient.

3.2 Multi-objective Particle Swarm Optimization (MOPSO)

Particle swarm optimization is used to solve multi-objective problems. Pareto dominance is a strategy the MOPSO algorithm employs to resolve multi-objective problems. A vector can present each particle i in the swarm $\boldsymbol{x_i} = \boldsymbol{x_1} + \boldsymbol{x_2} + \cdots + \boldsymbol{x_n}$, and their initial coordinates in the search space are chosen randomly where $\boldsymbol{n}$ is the number of particles present throughout the population. The particle's velocity can be expressed as $\boldsymbol{v_i} = \boldsymbol{v_1} + \boldsymbol{v_2} \dots . + \boldsymbol{v_n}$. The local best ($\boldsymbol{P_{lbest}}$) and the global best (($\boldsymbol{P_{gbest}}$) are the two criteria that the particle uses to determine its position when it begins to search. The archive contains the top local solutions. Hypercubes are created through these solutions. $\boldsymbol{P_{gbest}} = \boldsymbol{x_i}$ indicates the best particle position in the swarm based on the local best values that were previously stored. Equation (7) is used to update the particle velocity for each generation:

$$v_i(t+1) = \omega . v_i(t) + \varphi_1(P_{lbest} - x_i) + \varphi_2(P_{gbest} - x_i) \tag{7}$$

where $\boldsymbol{\varphi_1}$ and $\boldsymbol{\varphi_2}$ are random numbers between [0, 1] and $\boldsymbol{\omega}$ is the inertia weight. The position of the particle can be updated using equation (8).

$$x_i(t+1) = x_i(t) + v_i(t+1) \tag{8}$$

For each particle, the boundary condition is considered. A novel mutation operation is applied, to improve the search process. Then the repository updates to $\boldsymbol{P}_{gbest}$ and hypercubes are performed. The repository is updated by adding non-dominated solutions and removing dominated ones from the repository. The final step updates each particle's position by substituting the current best position ($\boldsymbol{P}_{lbest}$) for the previous best position ($\boldsymbol{P}_{gbest}$).

3.3 Dataset Description

The Kaggle repository provides free access to the PCOS dataset [24]. A total of 541 females contributed 41 input features and one output feature to this dataset. Of these 541 instances, 364 were unaffected individuals, and the remaining 177 were PCOS patients. These cases were split into two groups: 0 and 1, where 0 represents normal women, and 1 represents PCOS-positive women.

4. System Framework

This section explains the system design of the ANN-MOPSO-BP prediction model. Fig. 1 depicts the system architecture of the ANN-MOPSO-BP prediction model.

4.1 Preprocessing Subsystem

There were missing values in the PCOS dataset obtained from the Kaggle Repository. Table 1. provides information about the missing values in the PCOS dataset. In this study, the missing values were replaced by imputing the values that occurred the most frequently for the relevant class. The input for data normalization is a dataset with no missing values. Reduce the clinical dataset's irregular feature values into a given range by applying the min-max normalization method shown in equation (9).

$$Normalized(X) = \frac{E-E_{min}}{E_{max}-E_{min}}\left(E_{new\_max} - E_{new\_min}\right) + E_{new\_min} \tag{9}$$

where E is the feature value that must be normalized, X is the normalized feature value, $Emin$ is the minimum value of the feature, $Emax$ is its maximum value, $Enew\_max$ represents the maximum normalized value, and $Enew\_min$ represents the minimum normalized value.

4.2 Feature Selection

A machine learning model is created through the feature selection process, which involves minimizing the number of input features. In some situations, the effectiveness of a classifier model can be increased by lowering the number of input features, which would lower the cost of the modeling process. Machine learning classification problems can be solved using the correlation between the features. The association between each input feature and the class feature is statistically assessed using this method. The feature-class correlation and feature-feature inter-correlations were determined after each input feature, and the class is treated equally. The input features were selected based on the correlation value. Strong correlations were observed between the class variables and chosen input variables.

4.3 Training Subsystem

The training subsystem used ANN training with the MOPSO-BP algorithm. The ANN consists of three layers: namely, input, hidden layer and output layers. The weights between the input and hidden, hidden and output and biases are optimized using the MOPSO-BP algorithm. The MOPSO algorithm has two objectives: improving the diagnostic accuracy based on MSE and exploring the search space to overcome the disadvantage of stagnation. The ANN-MOPSO-BP classifier used a trained subsystem to diagnose PCOS.

4.4 Testing Subsystem

The testing subsystem used the ANN-MOPSO-BP classifier model to diagnose PCOS. The efficiency of the developed classifier is tested using the PCOS dataset obtained from the Kaggle machine learning repository. 70-30 and k-fold cross validation were used to train and test the classifier model. The testing subsystem uses preprocessed input features to the ANN-MOPSO-BP classifier to obtain diagnostic results.

Table 1 *Details about missing values*

Sl.No	Attribute Name	Number of missing values
1	Marriage Status (Yrs)	1
2	Fast food (Y/N)	1

5. Results and Discussions

This section presents the findings of the experiments and analyses. Table 2 displays the training parameters for the MOPSO-BP algorithm. The steps of the suggested ANN-MOPSO-BP classifier were displayed by displaying the results while taking into account the input PCOS dataset. The ANN-MOPSO-BP classifier's performance was assessed using 70-30 and k-fold cross-validation of training and test samples. Equations (10) and (12) are used to generate these performance metrics.

Table 2 *Parameter setting for the MOPSO algorithm*

Sl.no	Classifier	Parameters
1	NN-MOPSO-BP	$maximim\ generations = 250, learning\ rate\ \eta = 0.01, minimum\ error\ \varepsilon = 10^{-2}, number\ of\ particles = 50, \omega = 0.4, \varphi_1 and\ \varphi_2 = 0.9$

$$Sensitivity = \frac{TP}{(TP+FN)} \tag{10}$$

$$Specificity = \frac{TN}{(FP+TN)} \tag{11}$$

$$\text{Accuracy} = \frac{TP+TN}{(TP+FP+FN+TN)} \tag{12}$$

$$Positive\ Likelihood\ Ratio\ (PLR) = \frac{TP\ rate}{FP\ rate} = \frac{Sensitivity}{(1-Specificity)} \tag{13}$$

$$Negative\ Likelihood\ Ratio(NLR) = \frac{FN\ rate}{TN\ rate} = \frac{(1-Sensitivity)}{Specificity} \tag{14}$$

$$Positive\ Predictive\ Value(PPV) = \frac{TP}{(TP+FP)} \tag{15}$$

$$Negative\ Predictive\ Value(NPV) = \frac{TN}{(FN+TN)} \tag{16}$$

Figure 2 shows the features that were selected based on a 0.25 correlation threshold with the target value. The performance of the MOPSO-BP-trained ANN classifier is displayed in Figure 3. The red line represents the output for the target, and the blue line shows the output for the classifier. The best value obtained overall after each cycle is represented by the green line. The error value obtained for each sample is shown on the pink line. Based on the results, it can be said that CDSS systems can successfully use the ANN-MOPSO-BP classifier.

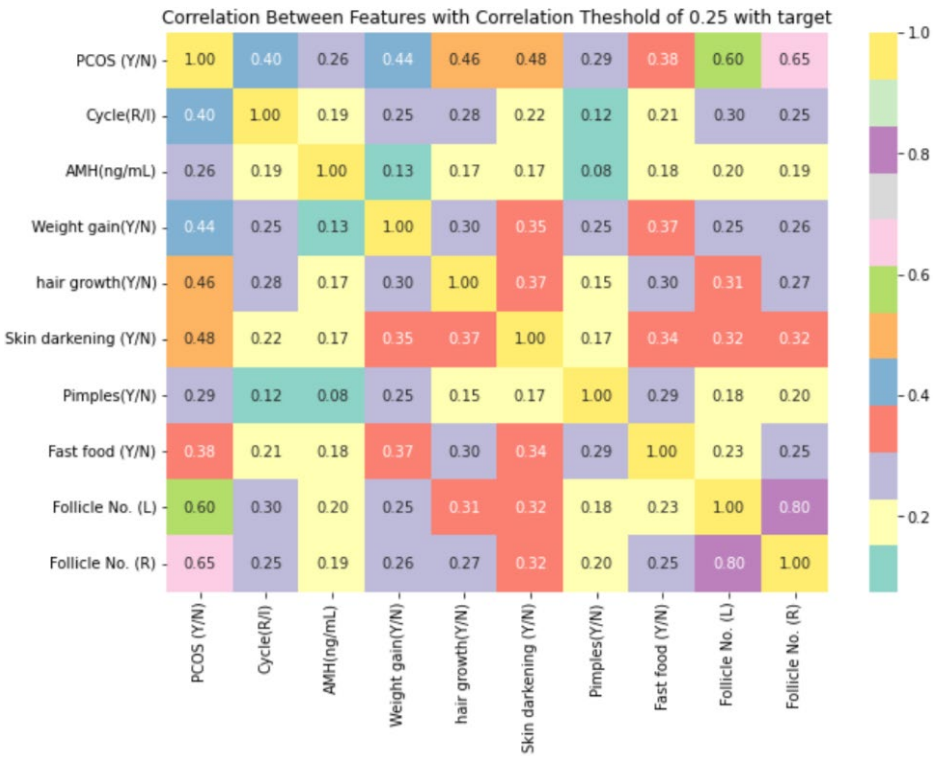

Fig. 2 *Correlation between selected features and target*

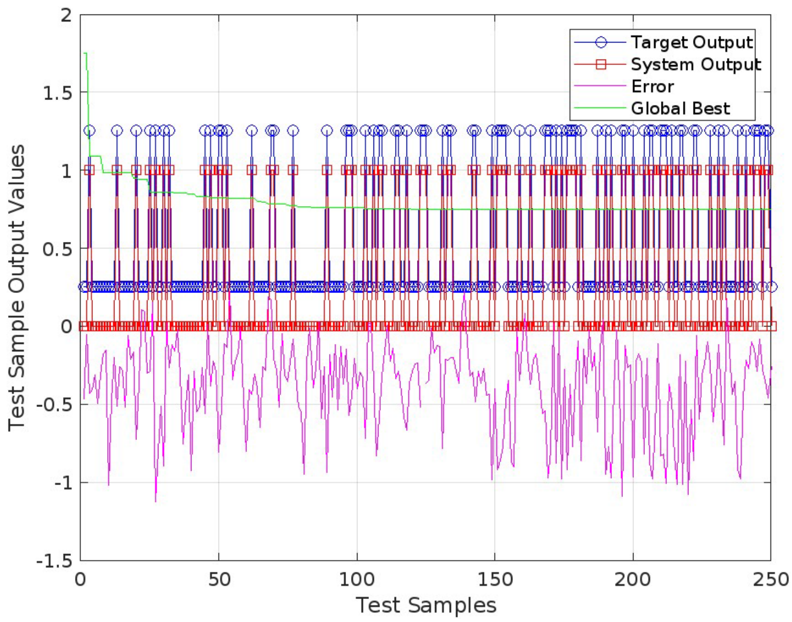

Fig. 3 *Performance of ANN-MOPSO-BP trained classifier output*

The ANN-MOPSO model displayed superior performance compared with both the ANN and ANN-PSO models. By integrating multi-objective optimization, it effectively balances various criteria such as minimizing the mean square error and preventing training stagnation. Consequently, this approach yields a more resilient model that exhibits better generalization capabilities for unseen data, as demonstrated by its holdout accuracy. Table 3 shows the accuracy results for three different classifiers using the holdout method and K-fold cross validation. Utilizing three classifiers.

Table 3 *Accuracies of three different classifiers using K-fold cross validation and holdout approach*

Classifier	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Fold 6	Fold 7	Mean Accuracy	Holdout Accuracy
ANN	0.666	0.666	0.675	0.675	0.675	0.675	0.675	0.67	70.55
ANN-PSO	0.910	0.897	0.831	0.844	0.883	0.909	0.896	0.87	90.18
ANN-MOPSO	0.884	0.910	0.909	0.844	0.896	0.883	0.870	0.88	92.02

The ANN Model exhibited the highest sensitivity and specificity among the four models, demonstrating its ability to effectively identify PCOS patients. The ANN-MOPSO Model showed a balanced performance, outperforming the ANN-PSO Model and demonstrating strong classification capabilities. The ANN Model had the highest positive and negative predictive values, indicating a high level of reliability for both positive and negative diagnoses. The ANN-MOPSO Model also performed well in terms of PPV and NPV, making it a strong candidate for use in clinical settings. The ANN-PSO Model showed lower PPV and NPV. The ANN Model demonstrated the highest positive likelihood ratio and the lowest negative likelihood ratio, indicating that it is highly effective in confirming the disease when the test is positive and in ruling out the disease when the test is negative. The ANN-MOPSO Model also performed well in terms of the PLR and NLR, whereas the ANN-PSO Model performed moderately well. The prevalence rates were similar among the models, suggesting a consistent representation of the dataset. The ANN model proved to be the most sensitive and specific for diagnosing PCOS, with high PPV and NPV, whereas the ANN-MOPSO model demonstrated strong overall performance. These findings support the use of ANN-MOPSO-based CDSS for PCOS diagnosis, providing gynecologists with reliable second-opinion resources. Future enhancements could focus on improving the sensitivity and computational efficiency and accommodating diverse patient datasets. Table 4 shows the diagnostic test evaluation of the PCOS dataset.

Table 4 *Diagnostic test evaluation of PCOS dataset using three classifiers*

Evaluation Metrics	ANN	95% Confidence Interval		ANN-PSO	95% Confidence Interval		ANN-MOPSO	95% Confidence Interval	
		From	To		From	To		From	To
Sensitivity (%)	91.30	79.21	97.58	80.36	67.57	89.77	87.27	53.29	93.32
Specificity (%)	97.20	92.02	99.42	91.59	84.63	96.08	94.44	88.30	97.93
PLR	32.57	10.63	99.73	9.55	5.04	18.09	12.11	5.45	26.93
NLR	0.09	0.04	0.23	0.21	0.13	0.37	0.35	0.24	0.51
Disease Prevalence (%)	30.07	22.93	38.00	34.36	27.11	42.19	33.74	26.53	41.55
PPV (%)	93.33	82.05	97.72	83.33	72.53	90.45	86.05	73.50	93.20
NPV (%)	96.30	91.06	98.52	89.91	83.95	93.82	85.00	79.46	89.25

The performance evaluations of various machine learning models for PCOS diagnosis revealed a range of accuracies, highlighting the strengths and limitations of each approach. Faris et al. (2022) achieved the highest accuracy of 99% with a K-M-SVM model, demonstrating the potential of combining clustering (K-means) and SVM for robust classification. This model excels in high-dimensional data spaces, making it particularly effective for complex datasets [17]. Danaei Mehr et al. (2022) also achieved a high accuracy of 98.89% using an Ensemble RF classifier, indicating that ensemble methods, which combine multiple learning algorithms, can significantly enhance prediction performance. This approach benefits from the diversity and complementary strengths of the individual classifiers [18].

Bhat (2021) obtained a notable accuracy of 95% using the CatBoost model, which is well known for handling categorical features. This model is advantageous for datasets with a mix of numerical and categorical data, as it provides reliable predictions without extensive preprocessing [21]. Tiwari et al. (2022) and Marreiros et al. (2022) achieved accuracies of approximately 93% using Random Forest models, which consistently performed well. These models are robust to overfitting and can efficiently handle large datasets. The use of ensemble techniques within random forests further enhances their reliability [16], [19]. Rakshitha and Naveen [15] achieved an accuracy of 89.03% with a hybrid approach combining SVM with a linear kernel and Logistic Regression. Although effective, this method may struggle with complex nonlinear relationships inherent in some clinical datasets [15].

Sreejith et al. (2022) also demonstrated a solid accuracy of 89.81% using a hybrid approach that integrates the Red Deer algorithm for feature selection with a Random Forest classifier. This approach benefits from optimized feature selection and enhances the overall model performance [23]. Table 5 compares the proposed method's accuracy to a few other approaches that have been published in the literature for the PCOS dataset.

Table 5 *Accuracy comparison of the proposed method to some previously published methods for the PCOS dataset*

Sl. no	Author	Method / Reference	Accuracy
1	Rakshitha and Naveen [15]	Hybrid SVM linear kernel with the LROp-	89.03%
2	Tiwai et al. [16]	RMSprop RF approach	93.25%
3	Faris et al. [17]	K-M-SVM	99%
4	Danaei Mehr et al. [18]	Ensemble RF classifier	98.89%
5	Marreiros et al. [19]	RF classifier	93.06%
6	Silva et al. [20]	Random Forest	86%
7	Bhat [21]	CatBoost and Extreme Boosting with Random Forest (XGBRF)	XGBRF – 89% CatBoost – 95%
8	Danaei Mehr & Polat [22]	Embedded feature selection with ensemble random forest classifier	98.89%
9	Sreejith et al. [23]	Red Deer algorithm for feature selection wrapper with random forest classifier	89.81%

6. Conclusion

This work has implemented a CDSS that employs an ANN-trained MOPSO-BP classifier model to determine if PCOS is present or absent in female patients. The CDSS model had an accuracy of 92.02%, 87.27% sensitivity, and 94.44% specificity. The classification outcomes demonstrated that the CDSS may employ the ANN-MOPSO-BP classifier to aid junior doctors in diagnosing PCOS. The results show that having a darker complexion, eating fast food, gaining weight, and growing more hair are all associated with a higher risk of PCOS. Therefore, the conditions above could also be referred to as PCOS symptoms. Patients are more likely to develop PCOS when their follicle counts are higher and their cycle periods are shorter. Furthermore, PCOS patients are typically 30 years old when they receive their diagnosis. This study can be furthered by creating CDSS to diagnose PCOS from the medical images obtained through various imaging modalities. Classifier models can be created utilizing the relevant features to predict the existence or absence of abnormalities from the medical images.

Acknowledgement

The authors would like to thank Chitkara University Research and Innovation Network, Chitkara University, Punjab, India, for their support.

Conflict of Interest

The authors declare no conflict of interest regarding the paper's publication.

Author Contribution

The author confirms sole responsibility for the following: study conception and design, data collection, analysis and interpretation of results, and manuscript preparation.

References

[1] Azziz, R., Carmina, E., Dewailly, D., Diamanti-Kandarakis, E., Escobar-Morreale, H. F., Futterweit, W., ... & Witchel, S. F. (2006). Criteria for defining polycystic ovary syndrome as a predominantly hyperandrogenic syndrome: an androgen excess society guideline. The Journal of Clinical Endocrinology & Metabolism, 91(11), 4237-4245.

[2] Ding, T., Hardiman, P. J., Petersen, I., Wang, F. F., Qu, F., & Baio, G. (2017). The prevalence of polycystic ovary syndrome in reproductive-aged women of different ethnicity: a systematic review and meta-analysis. Oncotarget, 8(56), 96351.

[3] Gupta, M., Singh, D., Toppo, M., Priya, A., Sethia, S., & Gupta, P. (2018). A cross sectional study of polycystic ovarian syndrome among young women in Bhopal, Central India. Int J Community Med Public Health, 5(1), 95-100.
[4] Joshi, B., Mukherjee, S., Patil, A., Purandare, A., Chauhan, S., & Vaidya, R. (2014). A cross-sectional study of polycystic ovarian syndrome among adolescent and young girls in Mumbai, India. Indian journal of endocrinology and metabolism, 18(3), 317-324.
[5] Mehreen, T. S., Ranjani, H., Kamalesh, R., Ram, U., Anjana, R. M., & Mohan, V. (2021). Prevalence of polycystic ovarian syndrome among adolescents and young women in India. Journal of Diabetology, 12(3), 319-325.
[6] Dunaif, A. (1992). Polycystic ovary syndrome. Current Issues in Endocrinology and Metabolism: Polycystic Ovary Syndrome, 377-384.
[7] Carmina, E. (2004). Diagnosis of polycystic ovary syndrome: from NIH criteria to ESHRE-ASRM guidelines. Minerva ginecologica, 56(1), 1-6.
[8] 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. Human reproduction, 33(9), 1602-1618.
[9] Richens, J. G., Lee, C. M., & Johri, S. (2020). Improving the accuracy of medical diagnosis with causal machine learning. Nature communications, 11(1), 3923.
[10] Rahman, M.M., Islam, A., Islam, F., Zaman, M., Islam, M.R., Sakib, M.S.A. and Babu, H.M.H., 2024. Empowering Early Detection: A Web-Based Machine Learning Approach for PCOS Prediction. Informatics in Medicine Unlocked, p.101500.
[11] Tiwari, S., Kane, L., Koundal, D., Jain, A., Alhudhaif, A., Polat, K., Zaguia, A., Alenezi, F. and Althubiti, S.A., 2022. SPOSDS: A smart Polycystic Ovary Syndrome diagnostic system using machine learning. Expert Systems with Applications, 203, p.117592.
[12] Suha, S.A. and Islam, M.N., 2023. Exploring the dominant features and data-driven detection of polycystic ovary syndrome through modified stacking ensemble machine learning technique. Heliyon, 9(3).
[13] Jain, P., Mishra, R.K., Deep, A. and Jain, N.K., 2024. Xplainable AI for deep learning model on PCOD analysis. In XAI Based Intelligent Systems for Society 5.0 (pp. 131-152). Elsevier.
[14] Coello, C. A., & Toscano Pulido, G. (2001, March). A micro-genetic algorithm for multiobjective optimization. In International conference on evolutionary multi-criterion optimization (pp. 126-140). Berlin, Heidelberg: Springer Berlin Heidelberg.
[15] Rakshitha, K. P., & Naveen, N. C. (2022). Op-RMSprop (optimized-root mean square propagation) classification for prediction of polycystic ovary syndrome (PCOS) using hybrid machine learning technique. International Journal of Advanced Computer Science and Applications, 13(6).
[16] Tiwari, S., Kane, L., Koundal, D., Jain, A., Alhudhaif, A., Polat, K., ... & Althubiti, S. A. (2022). SPOSDS: A smart Polycystic Ovary Syndrome diagnostic system using machine learning. Expert Systems with Applications, 203, 117592.
[17] Faris, N. N., & Miften, F. S. (2022). An intelligence model for detection of PCOS based on K-means coupled with LS-SVM. Concurrency and Computation: Practice and Experience, 34(21), e7139.
[18] Danaei Mehr, H., & Polat, H. (2022). Diagnosis of polycystic ovary syndrome through different machine learning and feature selection techniques. Health and Technology, 12(1), 137-150.
[19] Marreiros, M., Ferreira, D., Neto, C., Witarsyah, D., & Machado, J. (2022). Classification of polycystic ovary syndrome based on correlation weight using machine learning. In Big Data Analytics and Artificial Intelligence in the Healthcare Industry (pp. 150-176). IGI Global.
[20] Silva, I. S., Ferreira, C. N., Costa, L. B. X., Sóter, M. O., Carvalho, L. M. L., de C. Albuquerque, J., ... & Gomes, K. B. (2022). Polycystic ovary syndrome: clinical and laboratory variables related to new phenotypes using machine-learning models. Journal of Endocrinological Investigation, 1-9.
[21] Bhat, S. A. (2021). Detection of polycystic ovary syndrome using machine learning algorithms (Doctoral dissertation, Dublin, National College of Ireland).
[22] Danaei Mehr, H., & Polat, H. (2022). Diagnosis of polycystic ovary syndrome through different machine learning and feature selection techniques. Health and Technology, 12(1), 137-150.
[23] Sreejith, S., Nehemiah, H. K., & Kannan, A. (2022). A clinical decision support system for polycystic ovarian syndrome using red deer algorithm and random forest classifier. Healthcare Analytics, 2, 100102.
[24] Kottarathil Prasoon, (2019) Polycystic ovraian syndrome (PCOS) dataset retrieved, october 2022, https://www.kaggle.com/prasoonkottarathil/polycystic-ovary-syndrome-pcos.
[25] Alquhtani, S. A., & Muniasamy, A. (2022). Development of Effective Electronic Customer Relationship Management (ECRM) Model by the Applications of Web Intelligence Analytics. In Building a Brand Image Through Electronic Customer Relationship Management (pp. 44-63). IGI Global.
[26] Dahiya, N., Gupta, S., & Singh, S. (2022). A review paper on machine learning applications, advantages, and techniques. ECS Transactions, 107(1), 6137.

[27] Singh, A., & Ramkumar, K. R. (2021). Risk assessment for health insurance using equation modeling and machine learning. International Journal of Knowledge-based and Intelligent Engineering Systems, 25(2), 201-225.