Available online at
https://jcrinn.com/
https://crinn.conferencehunter.com/

e-ISSN : 2600-8793

Journal of Computing Research and Innovation 9(1) 2024

Journal of Computing Research and Innovation

Polycystic Ovary Syndrome (PCOS) Prediction System Using PSO-SVM

Lukman Hakim Shaufee[1], Hamidah Jantan[2]*, Ummu Fatihah Mohd Bahrin[3]

[1,2,3]*College of Computing, Informatics & Mathematics, Universiti Teknologi MARA Terengganu Branch, Kuala Terengganu Campus, 21080 Kuala Terengganu, Terengganu, Malaysia.*

ARTICLE INFO

Article history:
Received: 30 Jan 2024
Revised: 27 February 2024
Accepted: 29 February 2024
Online first: 1 March 2024
Published 1 March 2024

Keywords:
PCOS
PSO-SVM Model
Feature Selection
Improved SVM
Machine Learning

DOI:
10.24191/jcrinn.v9i1.414

ABSTRACT

A prevalent and complicated gynaecological condition that affects women's reproductive health is PCOS. However, delayed diagnosis and treatment are frequently caused by a lack of understanding of its signs and symptoms. To help users and specialized physicians identify and anticipate ovarian cysts early, a PCOS prediction system integrating PSO-SVM was created to solve this issue. This study explores the application of data mining techniques, using PSO-SVM, to predict PCOS in the field of gynaecology. The dataset was taken from the Kaggle benchmark dataset, owned by Karnika Kapoor. There are 42 selected features and attributes of the PCOS dataset. The system used Python-based data preprocessing, data splitting, and PSO-SVM optimization for predicting PCOS disease. The evaluation showed that PSO-SVM with 20 particles and 100 iterations achieved the best accuracy for feature selection with an accuracy of 90.18%. The system exhibited promising predictive abilities. To enhance accuracy and user experience, future work should focus on longitudinal data integration, expert decision support, and collaboration with medical experts. The developed PSO-SVM-based PCOS prediction system significantly improves risk assessment and early identification, aiding patients, and medical practitioners. It serves as a valuable decision support tool for doctors, enabling quick and accurate diagnosis for early intervention and specialized treatment plans.

1. INTRODUCTION

This study will discuss the application of data mining techniques, which is Particle Swarm Optimization with Support Vector Machine (PSO-SVM), in predicting Polycystic Ovary Syndrome (PCOS) in the field of gynaecology. PCOS is a condition where excessive androgen hormones lead to the development of ovarian cysts, affecting women's reproductive health. PCOS is a prevalent hormonal disorder in women of reproductive age and the symptoms may arise shortly after the onset of menstruation or in response to weight gain. It encompasses a range of issues, and not all women experience the same symptoms. Diagnosis is confirmed if a patient exhibits either a lack of ovulation (anovulation), elevated testosterone levels

[2*] Corresponding author. *E-mail address*: hamidahjtn@uitm.edu.my
https://doi.org/10.24191/jcrinn.v9i1

(hyperandrogenism), or the presence of multiple small cysts on the outer edge of the ovary (polycystic). The lack of awareness increases the risk of developing PCOS, which affects up to 18% of women of reproductive age and can have metabolic effects (Aggarwal & Pandey, 2020). Diagnosing PCOS is also challenging, involving multiple clinical tests and scans (Denny et al., 2019). Many women, especially those trying to conceive, may not realize they have it, which can affect fertility. Community awareness and healthcare provider knowledge are essential (Soni & Vashisht, 2018). Early detection of this PCOS is crucial for effective management and improved reproductive health outcomes. This study proposes an approach using simple yet effective clinical and metabolic criteria for early PCOS identification.

Machine Learning (ML) algorithms have emerged as promising tools for predictive healthcare analytics, offering the potential to enhance PCOS prediction. Numerous studies have explored the application of ML, particularly ML algorithms like SVM (Denny et al., 2019; Vedpathak & Thakre, 2020), Random Forest (Vedpathak & Thakre, 2020), and artificial neural networks (Meena et al., 2015), in predicting PCOS. For example, a study by Sheikdavood and Bala (2023) applied Adaptive K-means with Reptile Search Algorithm for cyst segmentation, demonstrating exceptional precision and effectiveness in PCOS diagnosis. Research by Faridoon et al. (2023) implemented a Genetic Algorithm (GA) for feature selection in PCOS prediction, achieving notable accuracy rates. The scholars use GA to extract features, and a logistic regression model is used for the classification. All the results of the models are convincing and most of the results outperform form the baseline model.

The hybrid of the PSO-SVM algorithm could enhance accuracy in predicting PCOS, which has been successful in other domains. This study intends to assist in further diagnosis, reduce treatment duration, and empower patients to take early preventive actions. It also aims to create a prediction system to assist OB-GYN specialists and patients in identifying PCOS based on symptoms, potentially reducing the complexity of diagnosis, and improving reproductive health outcomes. The research aims to develop and evaluate a PCOS prediction system using the PSO-SVM algorithm. As for the project scope, there are three different items to be highlighted, the data using benchmark data from 2019-2022 from Kaggle by Karnika Kapoor, the targeted users would be Gynecologists, and medical assistants the female patients, the method and prototype development will be using SVM and enhance the algorithm performance with PSO. This project benefits the community by helping women detect PCOS early stage and aiding quicker diagnosis for gynaecologists. It also advances PCOS treatment in the gynaecology field. For developers, it showcases the use of PSO to enhance prediction systems and can inspire future improvements. Additionally, it can serve as a reference for students and other researchers working on PCOS prediction and optimization techniques.

Therefore, this study will be using SVM as an algorithm and further optimized through the integration of PSO. The implementation will be carried out using Kaggle’s readily accessible benchmark dataset, simplifying the development, training, and testing phases of the prediction system. Beyond its practical applications, this project serves as a reference for students and researchers working on PCOS prediction and optimization techniques, showcasing the potential of PSO in enhancing prediction systems and inspiring future advancements.

2. RELATED WORK

2.1 Algorithms and Techniques in Predicting PCOS

The gynaecology domain employs various methods and approaches for diagnosing PCOS. These include conducting a thorough physical examination, ultrasonography of the ovaries and pelvis, blood samples, and detailed family history. However, diagnosing PCOS can be challenging due to the vagueness of symptoms and the complexity of the disease. The accuracy of the diagnosis and the diagnosing process

https://doi.org/10.24191/jcrinn.v9i1

can be left unsatisfied, frustrated, and puzzled. Table 1 summarizes the related studies about the prediction of PCOS using various algorithms and techniques.

Table 1. PCOS Prediction using various machine learning and hybrid algorithms

Related Works	Dataset	Algorithm/Technique	Result
(Vedpathak & Thakre, 2020)	Kaggle benchmark dataset, owned by Prasoon Kottarathil	Random Forest, SVM, Logistic Regression, Gaussian Naïve Bayes, KNN Classifier	Random Forest Classifier is highlighted as the most reliable and accurate among the compared algorithms, with an accuracy of 90.9%
(Denny et al., 2019)	541 cases collected from various infertility treatment centers at Thrissur, comprising women in the reproductive age group (18-40 years)	Principal Component Analysis (PCA) Naïve Bayes, Logistic Regression, K-Nearest Neighbor (KNN), Classification and Regression Trees (CART) Random Forest Classifier, SVM	Identified as the most suitable and accurate method for PCOS prediction with an accuracy of 89.02%.
(Bharati et al., 2020)	UCI Machine Learning Repository consist of 541 records of women, with labelled Normal and PCOS	Gradient boosting, random forest, logistic regression, and hybrid random forest and logistic regression (RFLR)	The best 10 features are good enough to predict PCOS disease, and RFLR exhibits the best testing accuracy of 91.01% and recall value of 90% using 40-fold cross-validation.
(Faridoon et al., 2023)	A dataset containing the clinical and biochemical characteristics of 109 patients, including 36 with PCOS	GA to extract features, and a logistic regression model is used for the classification	The proposed model outperformed the baseline model that used all features, which had an accuracy of 86.2%, and produced a markedly improved accuracy of 95.4%.
(Sheikdavood & Bala, 2023)	Kaggle repository and a dataset prepared by Prasoon Kottarathil	Adaptive K-means with Reptile Search Algorithm (IAKmeans-RSA) for cyst segmentation and Deep Neural Network (DNN) for PCOS classification	Classification achieved an accuracy of 99.1%, precision of 98.3%, recall of 98.8%, F1-score of 98.5%, and ROC-AUC of 0.9.

In summary, various datasets, and algorithms, including traditional machine learning (ML) and hybrid approaches, were employed for the prediction of PCOS. The datasets encompassed a Kaggle benchmark dataset, a dataset from various infertility treatment centers, and a dataset from the UCI Machine Learning Repository. Traditional ML algorithms such as Random Forest, SVM, Logistic Regression, Gaussian Naïve Bayes, KNN Classifier, Principal Component Analysis (PCA), K-Nearest Neighbor (KNN), and Classification and Regression Trees (CART) were utilized. Additionally, hybrid approaches combining Adaptive K-means with Reptile Search Algorithm (IAKmeans-RSA) for cyst segmentation and Deep Neural Network (DNN) for PCOS classification were employed. These algorithms successfully predict the presence of PCOS with high accuracy, precision, recall, and F1-score, demonstrating their potential for effective early detection and diagnosis.

2.2 PSO-SVM in Various Domains

Particle Swarm Optimization with Support Vector Machine (PSO-SVM) represents a potent synergy between two intelligent algorithms that has garnered significant attention in various domains due to its remarkable capabilities (He & Song, 2014). PSO, inspired by the collective behaviours of birds and fish, collaborates with SVM, a robust machine learning technique, to bring about enhanced performance, accuracy, and efficiency in solving complex optimization and prediction tasks. Table 2 shows the implementation of PSO-SVM in various domains showcasing its efficacy in enhancing accuracy, stability, and efficiency in different application areas.

https://doi.org/10.24191/jcrinn.v9i1

Table 2. PSO-SVM in various domains/applications

Related Works	Algorithm/Technique	Domain	Result
(Hitam et al., 2019)	Particle Swarm Optimization with Support Vector Machine (PSO-SVM)	Finance and Economics Domain	SVM-PSO outperformed another classifier with an accuracy of 97%.
(Xue & Jieru, 2022)		Medical, and Biological Domain	SVM features are made more sensible after PSO optimization and the model's accuracy improves.
(Tang, 2022)		Biometrics Domain	The accuracy of prediction and detection efficiency simultaneously increase.
(Nugraha et al., 2019)		Computer Science and Informatics Engineering	Using PSO-SVM, the accuracy enhancement was accomplished.
(Liao et al., 2019)		Electrical Engineering and Energy Domain	With PSO-SVM, the load forecasting model and load early warning model have better accuracy and stability.

Based on the studies mentioned, it can be concluded that the application of PSO to enhance the performance of SVM has shown promising results in various domains. The use of PSO-SVM has led to improved accuracy in cryptocurrency price forecasting, heart disease detection, face recognition, journal rank classification, and short-term load forecasting for charging stations. These findings highlight the effectiveness of PSO in optimizing SVM algorithms for different applications, leading to improved prediction accuracy, reduced model training time, and enhanced classification precision and generalization. Overall, the combination of PSO and SVM has demonstrated its potential in addressing complex optimization and prediction tasks across diverse domains.

To address this issue, a PCOS prediction system based on the SVM-PSO algorithm is suggested in this study. This system is hoping can lessen and avoid human error that could affect PCOS diagnosis and aid in the early management of symptoms. The specialist can utilize this system as a supplemental tool to expedite the diagnosis of PCOS and save time while examining patients.

3. RESEARCH METHOD

As for the research method for this project, it is some phases to be applied to it, which are data acquisition, algorithm design, and integration of PSO-SVM. As for the data acquisition, the data source will be selected for this project. The secondary data source which are benchmark dataset from the Kaggle site provided by Karnika Kapoor in 2021. The dataset was retrieved on 3rd November 2022. As for the retrieved dataset, it consists of patient details without infertility. As for the data description of the PCOS diagnosis dataset, the dataset provides all physical and clinical features needed to diagnose PCOS and difficulties linked to infertility. According to each dataset, there are 42 selected features for PCOS without infertility dataset. The dataset has 541 instances altogether.

The sample of clean data will be visualized in this section as it relates to data representation. Data representation can be created after the data preparation procedure. Multiple pre-processing data preparation steps have been taken to achieve a clean dataset, including the removal of duplicate instances, the replacement of missing values, and attribute selection for finalizing features as in Table 3.

https://doi.org/10.24191/jcrinn.v9i1

Table 3. Data representation

No.	Attribute	Type	Unit/Measurement
1	Age	Numeric	years
2	Height	Numeric	cm
3	BMI	Numeric	-
4	Pulse Rate	Numeric	bpm
5	Haemoglobin (Hb)	Numeric	g/dl
6	Cycle	Numeric	R/I
7	Pregnant	Nominal	Y/N
8	FSH/LH	Numeric	-
9	Waist: Hip Ratio	Numeric	-
10	TSH	Numeric	mIU/L
11	AMH	Numeric	ng/mL
12	Vit D3	Numeric	ng/mL
13	PRG	Numeric	ng/mL
14	Hair Loss	Nominal	Y/N
15	Fast Food	Nominal	Y/N
16	BP Systolic	Numeric	mmHg
17	BP Diastolic	Numeric	mmHg
18	Follicle No. (R)	Numeric	-
19	Avg. Follicle Size (L)	Numeric	mm
20	Avg. Follicle Size (R)	Numeric	mm

Regarding algorithm design, the flowchart essentially illustrates the operation of the SVM-PSO model through a diagrammatic representation of the model's sequential flow such like in Fig. 1.

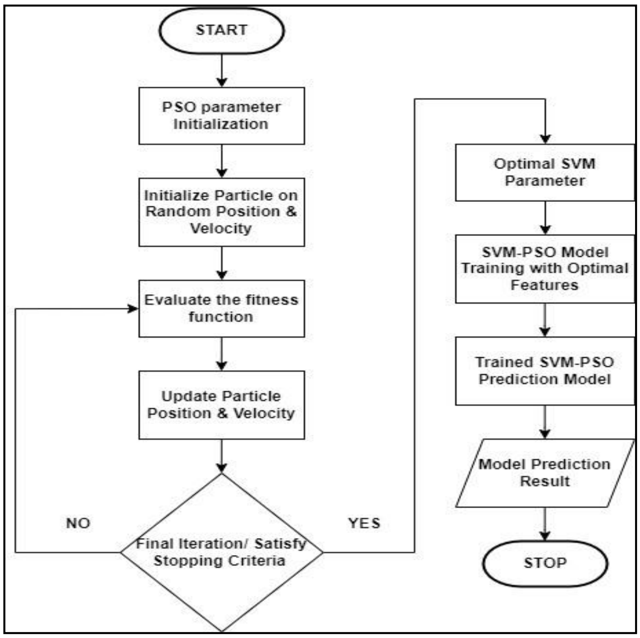

Fig. 1. PSO-SVM hybrid algorithm flowchart

https://doi.org/10.24191/jcrinn.v9i1

For the PSO-SVM algorithm flowchart, the flow starts with PSO features initialization together with initialising particles on random position and velocity (Thomas & Kavitha, 2020). In the modelling, this involves setting up the initial particles and PSO features for the SVM-PSO algorithm, followed by configuring PSO settings such as population size, iteration limits, inertia weight, and acceleration coefficients. Subsequently, the fitness function is evaluated, and particles are updated based on the fitness analysis. The update includes adjusting the velocity and position of each particle. Once the maximum number of iterations is reached, the evolutionary process concludes, and the SVM model training phase begins. In the SVM model training phase, features are determined, and the values of ideal features and parameters are acquired upon reaching the specified number of iterations. After completing the training and verification process, the PSO-SVM model is constructed (Han & Bian, 2018). This entire process is illustrated in Figure 1 as a flow diagram of the PSO-SVM model, providing a visual representation of the algorithm's stages.

Additionally, the steps involved in standardized data preparation, PSO objective function, SVM classifier implementation, swarm initialization, and optimization. Following the completion of the initialization process, the features will have the fitness function evaluated. The particle position and velocity will be modified to reflect the outcome in the best possible way. The stopping condition would then be either satisfied or not in an if statement (Sanyal, 2023). The iteration will continue toward the fitness function if the stopping requirements are not met. If the stopping criteria have been satisfied, the optimal SVM features have been required and the features will go into PSO-SVM model training with optimal features selected to achieve a trained PSO-SVM prediction model (Sanyal, 2023). The model prediction can be used once the PSO-SVM prediction model has been trained with the optimal features. The PSO-SVM model execution will end once the output has been generated.

Fig. 2 above illustrates several significant processes for the examination of the conceptual framework. Start by acquiring the data and pre-processing the data before it goes for the feature selection process by PSO. In PSO, the process of initialization of population particles, an objective function which evaluates the fitness, updating on personal best (pBest) and global best (gBest), together with the position of each particle, then it will check on the termination condition (Li et al., 2020). Once the selected features have been acquired it will go for data splitting for training and testing data to be trained by SVM. The SVM algorithm will go through four stages when the input has been prompted, the first stage begins with the initiation of a hyperparameter of learning rate, number of iterations and lambda parameters. The data are then fitted to the input data, X and target label, Y in the following stage. Continue to the following stage, which involves initializing the bias value (self. b) and the weight vector (self. w), and subsequently updating the weight, specifically in the gradient descent. The trained model of PSO-SVM has been acquired and is ready to use with the test dataset and user input prompt. The SVM prediction result is ready to be displayed to the user once it has passed through all four stages. These describe the conceptual framework’s flow as it relates to this prediction system.

https://doi.org/10.24191/jcrinn.v9i1

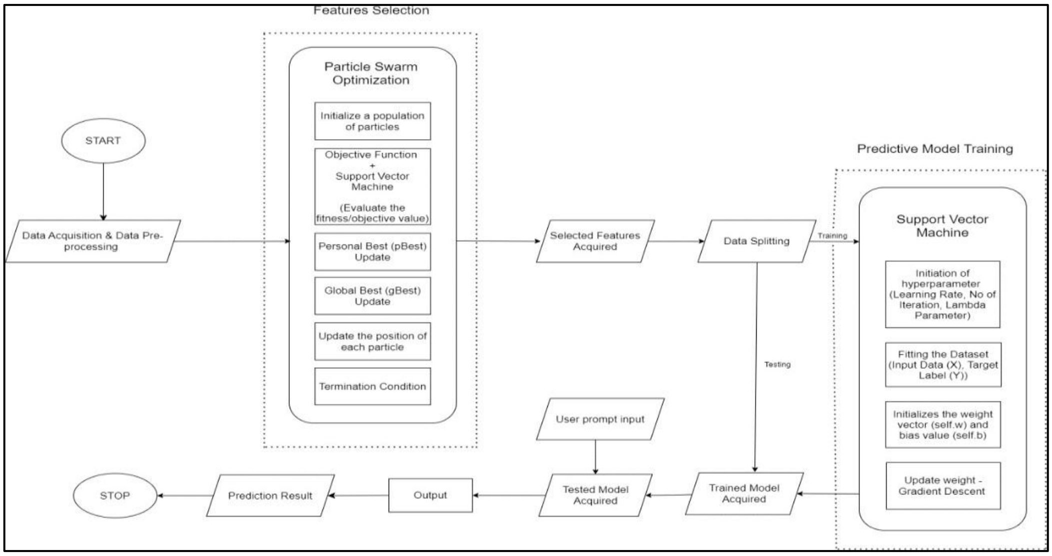

Fig. 2. Integration PSO-SVM Predictive System

4. RESULT ANALYSIS

This evaluation study examines the performance of three different approaches to improve prediction accuracy and optimize machine learning algorithms, specifically SVM. There are three (3) different experiments that evaluated which are an experiment with a basic single SVM algorithm, an experiment with PSO-SVM, and together with a PSO variation with modified particles and iterations. As for the accuracy of each experiment evaluation result, is taken by using confusion matric evaluation libraries.

The first experiment involves a single SVM without feature selection for both training and testing. This experiment includes testing on three distinct training-testing split ratios: 70:30, 80:20, and 90:10. Additionally, three types of SVM kernels, namely Linear, Polynomial, and RBF, are employed. For the 70:30 training-testing split ratio, the Polynomial SVM kernel yields the lowest accuracy, while the Linear SVM kernel achieves the highest accuracy at 88.96%. Shifting to the 80:20 training-testing split ratio, the Linear SVM kernel records the lowest accuracy at 85.32%, whereas the RBF SVM kernel attains the highest accuracy at 87.56%. In the last split ratio, which is 90:10, Polynomial and RBF had the same accuracy which is 85.45% which is slightly higher than the SVM Linear kernel with 83.64%. Table 4 shows the experiment setup for experiment 1 and Fig. 3 shows the accuracy of tested model.

Table 4. Experiment 1: Basic SVM Algorithm

Training Testing	Accuracy Percentage (%)		
Split Ratio/ SVM Kernel	**Linear**	**Polynomial**	**RBF**
70:30	88.96	85.28	87.12
80:20	85.32	86.24	87.56
90:10	83.64	85.45	85.45

https://doi.org/10.24191/jcrinn.v9i1

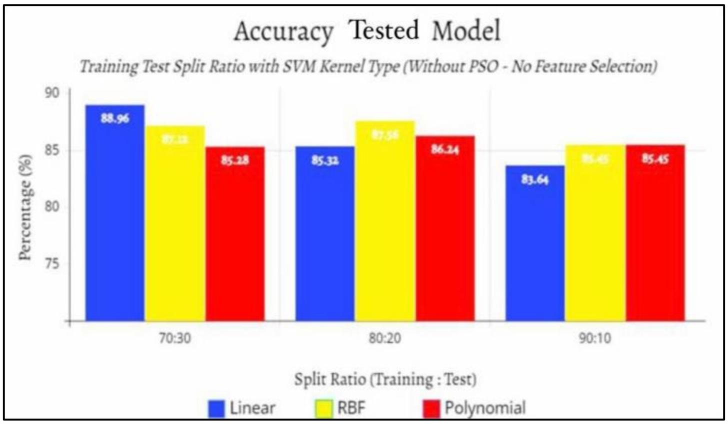

Fig. 3. Comparison of algorithms performance without PSO - Feature Selection

The experiment would be on the hybrid algorithm, in which to be specific the PSO-SVM hybrid algorithm where the features selection happened in PSO process flow and train and tested the model with selected features only. The experiment is the same as the previous experiment of single SVM but with the addition of optimization. There are three different split ratios together with three different SVM kernel which is 70:30, 80:20, 90:10, and Linear kernel, Polynomial kernel, and RBF kernel, respectively. The number of particle and iteration of PSO for optimization have been set to 20 particles with 100 iterations, where the detailed explanation is on the next experiment discussion.

In this experiment, as for the 70:30 split ratio, the accuracy for Linear and RBF kernel of SVM improved rather than the single SVM algorithm, from 88.96% to 90.79% for Linear kernel and from 87.12% to 91.41%. The improvement on Polynomial kernel is just a slight improve of 0.61%. As for the 80:20 split ratio, the Linear kernel and RBF kernel had a massive gap improvement from 85.32% to 93.66%, and 87.56% to 93.57, respectively. Yet, for the Polynomial kernel type, there is no big gap of improvement, the increment just goes up by 0.92%. The last split ratio which is 90:10, all these three SVM kernel types had a great improvement in accuracy as shown in Table 5. As for the experiment evaluation results, SVM Polynomial kernel had a slight improvement for each split ratio, while RBF shows a consistent improvement of accuracy by above 90% yet the highest accuracy among all the split ratio tests within all SVM kernel types, it goes to SVM Linear kernel with 80:20 training testing split ration with 93.66% of accuracy.

Table 5. Experiment 2: Hybrid PSO-SVM algorithm

Training Testing	Accuracy Percentage (%)		
Split Ratio/ SVM Kernel	**Linear**	**Polynomial**	**RBF**
70:30	90.79	85.89	91.41
80:20	93.66	87.16	93.57
90:10	89.09	89.10	92.36

https://doi.org/10.24191/jcrinn.v9i1

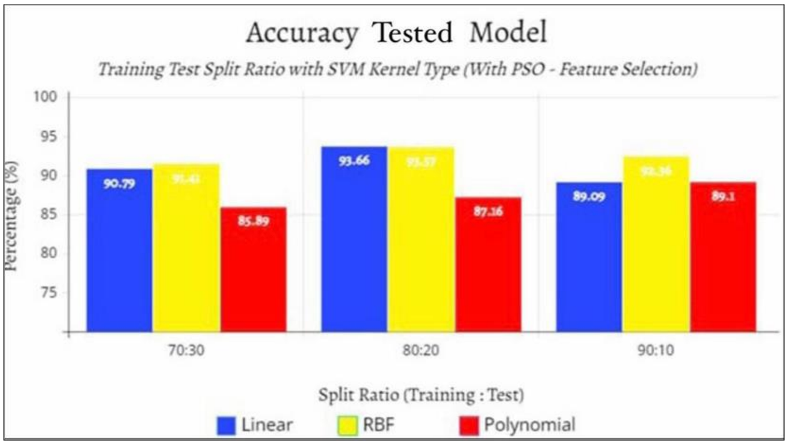

Figure 4. Comparison of Algorithms Performance with PSO - Feature Selection

The comparison of these two aims to determine which would be more suited for PCOS prediction. The variations in PCOS prediction between the SVM model and the PSO-SVM model are notable in this analysis. SVM locates a line that divides data into distinct groups, whereas PSO-SVM improves the accuracy of SVM by optimizing its parameters. Based on Fig. 5, the comparison between SVM and PSO-SVM models using Linear kernel type. Regarding observations, when employing a split ratio of 70:30, the PSO-SVM model demonstrates a slightly higher accuracy compared to the SVM model, showing an increment of 1.83% accuracy. Moving to an 80:20 split ratio, the improvement in the PSO-SVM model becomes more pronounced, surpassing the accuracy of the individual SVM model with an increment of 8.34%. However, at a split ratio of 90:10, both model accuracies decrease, yet the PSO-SVM model still outperforms the standalone SVM model.

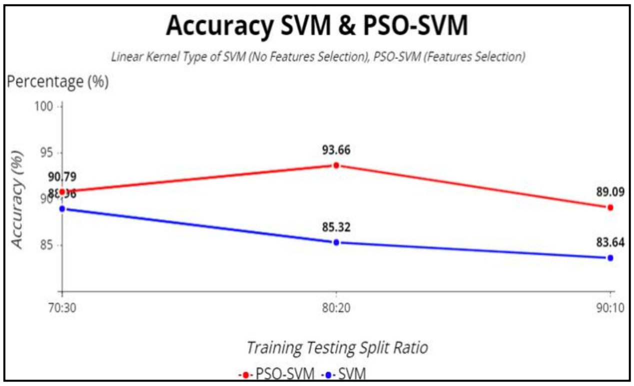

Fig. 5. Comparison of linear SVM & PSO-SVM model

By the visualization of Fig. 6, the observation can be seen that the increment of accuracy for the PSO-SVM model for all three split ratios had an increment higher than the single SVM model. The accuracy increments for 70:30, 80:20, and 90:10 is 4.29%, 6.01%, and 6.91%, respectively. This shows that the accuracy for both models is different, and for the split ratio 90:10, both had a decrease compared to the 80:20 split ratio.

https://doi.org/10.24191/jcrinn.v9i1

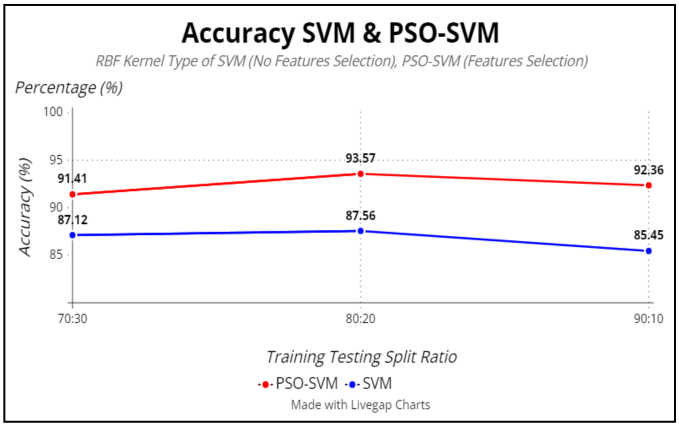

Fig. 6. Comparison of RBF SVM & PSO-SVM model

Based on Fig. 7, the observation of increment of accuracy for the PSO-SVM model can be made. On both 70:30 and 80:20 split ratios, both had a linear increment of accuracy, with increments of 0.61% and 0.92%, respectively. On the other hand, for a 90:10 ratio split, the PSO-SVM model keeps on improving while the single SVM model goes down. The difference in accuracy between these two models for a 90:10 split ratio is 3.65%.

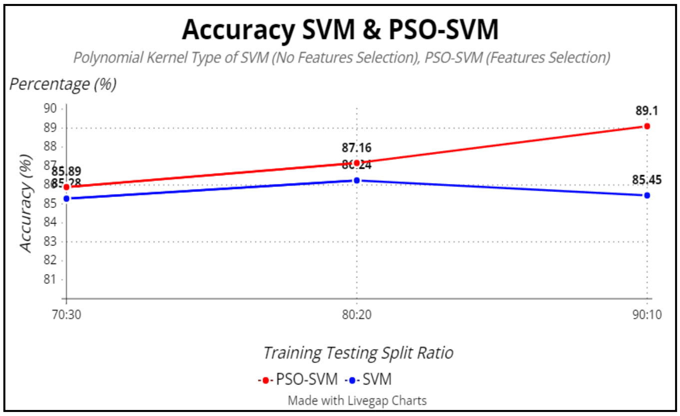

Fig. 7. Comparison of Polynomial SVM & PSO-SVM model

The impact of the data for prediction in SVM depends on the kernel selection. Radial basis function (RBF), kernels are best suited for data that can be separated into linear patterns, while polynomial kernels are best suited for complex relationships. Linear kernels are used for PCOS prediction because of their interpretability, insights into feature relevance, and applicability for linearly separable data (Hansen et al., 2011). These aid medical practitioners in comprehending the model's choices and locating crucial elements in the PCOS diagnosis. However, the choice of kernel depends on the properties of the data, nonlinearity may favour RBF or Polynomial kernel, needing testing and cross-validation.

In Table 6, the table itself consists of the number of particles that occur in PSO giving the best percentage of accuracy. The tweak of these number of particles together with a default number of iterations, gives the different output of accuracy on the best features. With 20 particles, iteration of 100 gave the best accuracy of feature selection with 90.18%. The same goes for several particles of 30 with the iteration of

https://doi.org/10.24191/jcrinn.v9i1

100, the accuracy achieved for the best feature selection on that number of particles is 89.57%. As for several particles 50, it achieves 87.12% accuracy on 100 iterations. Meanwhile, the default value of particles which is 80, achieves lower among all these three selected ranges of particles with 79.65% accuracy. Among all the evaluations within the range of several particles, the conclusion can be made that the number of particles of 20 together with iteration of 100 is the best setting for feature selection on selecting the best features for predicting PCOS disease.

Table 6. Accuracy Percentage (%) in Various Iterations

No. of Particle/ Iteration	**Accuracy Percentage (%)**
	100
20	90.18
30	89.57
50	87.12
80 (Default Value)	79.65

All experiments reveal a significant relationship that highlights the performance of the SVM algorithm. The experiments also shed light on the handling of PSO settings to optimize feature selection, contributing to the enhancement of the SVM algorithm through PSO optimization.

5. DISCUSSION

The result analysis of the proposed PSO-based approach for feature selection in predicting PCOS is compared with benchmark studies using several algorithms. With 20 particles and 100 iterations, the PSO approach achieved the highest accuracy at 90.18%, outperforming configurations with 30, 50, and the default 80 particles. In comparison to benchmark studies, the PSO-based method falls within the reported accuracy range of established algorithms such as Random Forest (RF) and Principal Component Analysis (PCA), conducted by Denny et al. (2019). The result is slightly lower than ours, with 89.02% accuracy. However, some benchmark studies, such as the one using DNN, demonstrate higher accuracies (Sheikdavood & Bala, 2023). The results showcase remarkable performance achieving an outstanding accuracy of 99.1%. Since there are slightly different accuracies in those experiments, it is important to consider other factors like computational efficiency, interpretability, and generalizability when choosing an approach (Shareef & Yosefi, 2021).

On the other hand, in the medical field, a study by Wang et al. (2021), a hybrid PSO-SVM is used to distinguish between the brains of Alzheimer's disease (AD) patients and an age-matched normal control group. The result showing simulation results showed that the highest classification accuracy of the feature combination determined by PSO was 97.22%, slightly higher than our study. Another study by Shuran and Yian (2020) presents an improved PSO-SVM based on Gray Relational Analysis, achieving 95.65% accuracy in distinguishing brains using the SVM-PSO model. This study also shows a great result compared to ours. The variations in outcomes could be attributed to distinctions in dataset characteristics (Oreški et al., 2017) and methodologies employed for feature selection. From the presented information, it can be concluded that the PSO-SVM approach successfully demonstrates achieving competitive accuracy rates.

6. CONCLUSION

In conclusion, the comprehensive evaluation study investigated three distinct approaches to enhance prediction accuracy and optimize machine learning algorithms, specifically focusing on the SVM. The experiments encompassed a basic single SVM algorithm, the hybrid PSO-SVM algorithm, and a modified PSO variation with adjusted particles and iterations. The evaluation utilized confusion matrix evaluation libraries to assess accuracy. The results emphasize the promising role of the PSO-SVM hybrid algorithm

https://doi.org/10.24191/jcrinn.v9i1

in improving PCOS prediction accuracy. Future iterations should focus on evolving data, expert guidance, and collaboration with medical specialists to overcome these limitations and further advance PCOS prediction. Although the algorithm enhances the precision and effectiveness of PCOS diagnosis, the limitations exist in this study. The system's dependency on accurate and comprehensive user input data, comprising biodata, reproductive parameters, and medical details, introduces the risk of inaccurate or incomplete data affecting prediction accuracy. Implementing data validation and verification measures, including user prompts to confirm data accuracy and informative error messages for blank or conflicting input fields, could improve the reliability of PCOS risk assessments and provide more insightful information to users. In summary, the study's objectives were achieved by uncovering the significance of combining PSO with a SVM for precise PCOS prediction. The proposed PCOS prediction system utilizing PSO-SVM considers various factors and exhibits promise, despite certain limitations. The findings contribute to the field of PCOS prediction, offering insights to both patients and medical practitioners, while acknowledging the need for continuous improvement and collaboration.

7. ACKNOWLEDGEMENTS/FUNDING

The authors would like to thank everyone, just everyone!

8. CONFLICT OF INTEREST DISCLOSURE

We declare that there is no conflict of interest regarding the publication of this article.

9. REFERENCES

Aggarwal, S., & Pandey, K. (2020). An analysis of PCOS disease prediction model using machine learning classification algorithms. *Recent Patents on Engineering, 15*(6), 53-63. https://doi.org/10.2174/1872212115999201224130204

Bharati, S., Podder, P., & Mondal, M. R. H. (2020, 5-7 June 2020). Diagnosis of polycystic ovary syndrome using machine learning algorithms. *2020 IEEE Region 10 Symposium (TENSYMP)*. IEEE. https://10.1109/TENSYMP50017.2020.9230932

Denny, A., Raj, A., Ashok, A., Ram, C. M., & George, R. (2019, 17-20 Oct. 2019). i-HOPE: Detection and prediction system for polycystic ovary syndrome (PCOS) using machine learning techniques. *TENCON 2019 - 2019 IEEE Region 10 Conference (TENCON)*. IEEE. https://10.1109/TENCON.2019.8929674

Faridoon, F., Ali, R. H., Abideen, Z. U., Shahzadi, N., Ijaz, A. Z., Arshad, U., Ali, N., Imad, M., & Nabi, S. (2023, 9-10 Oct. 2023). Prediction of polycystic ovary syndrome using genetic algorithm-driven feature selection. *2023 International Conference on IT and Industrial Technologies (ICIT)*. IEEE. https://10.1109/ICIT59216.2023.10335879

Han, B., & Bian, X. (2018). A hybrid PSO-SVM-based model for determination of oil recovery factor in the low-permeability reservoir. *Petroleum*, *4*(1), 43-49.

Hansen, K., Baehrens, D., Schroeter, T., Rupp , M., & Müller , K. R. (2011). Visual interpretation of kernel-based prediction models. *Molecular Informatics*, *30*. https://doi.org/10.1002/minf.201100059

He, Y., & Song, C. (2014). A novel particle swarm optimization algorithm based on iterative chaotic map with infinite collapses for global optimization. International Joint Conference on Computational Intelligence (pp. 38-48).

Hitam, N. A., Ismail, A. R., & Saeed, F. (2019). An optimized Support Vector Machine (SVM) based on Particle Swarm Optimization (PSO) for cryptocurrency forecasting. *Procedia Computer Science*, *163*, 427-433. https://doi.org/10.1016/j.procs.2019.12.125

Li, X., Wu, S., Li, X., Yuan, H., & Zhao, D. (2020). Particle swarm optimization-support vector machine model for machinery fault diagnoses in high-voltage circuit breakers. *Chinese Journal of Mechanical Engineering*, *33*(1), 6. https://doi.org/10.1186/s10033-019-0428-5

Liao, X., Kang, X., Li, M., & Cao, N. (2019, 12-13 Jan. 2019). Short term load forecasting and early warning of charging station based on PSO-SVM. *2019 International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS)*. IEEE. https://10.1109/ICITBS.2019.00080

Meena, K., Manimekalai, M., & Rethinavalli, S. (2015). Correlation of Artificial Neural Network classification and NFRS attribute filtering algorithm for PCOS data. *Int. J. Res. Eng. Technol*, *4*(3), 519-524.

Nugraha, Y. R., Wibawa, A. P., & Zaeni, I. A. E. (2019). Particle swarm optimization – Support vector machine (PSO-SVM) algorithm for journal rank classification. *2019 2nd International Conference of Computer and Informatics Engineering (IC2IE)* (pp. 69-73).

Oreški, D., Oreski, S., & Klicek, B. (2017). Effects of dataset characteristics on the performance of feature selection techniques. *Appl. Soft Comput.*, *52*, 109-119.

Sanyal, S. (2023, August 9th, 2023). *An Introduction to Particle Swarm Optimization (PSO Algorithm)*. https://www.analyticsvidhya.com/blog/2021/10/an-introduction-to-particle-swarm-optimization-algorithm/

Shareef, D. M. A. M., & Yosefi, G. A. (2021). A literature review of feature selection methods. *EasyChair Preprint*.

Sheikdavood, K., & Bala, M. P. (2023). Polycystic ovary cyst segmentation using adaptive k-means with reptile search algorith. *Information Technology and Control*, *52*(1), 85-99.

Shuran, C., & Yian, L. (2020). Breast cancer diagnosis and prediction model based on improved PSO-SVM based on gray relational analysis. *2020 19th International Symposium on Distributed Computing and Applications for Business Engineering and Science (DCABES)*. IEEE. http://10.1109/DCABES50732.2020.00067

Soni, P., & Vashisht, S. (2018). Exploration on polycystic ovarian syndrome and data mining techniques. *2018 3rd International Conference on Communication and Electronics Systems (ICCES)*. IEEE. https://doi.org/10.1109/CESYS.2018.8724087

Tang, W. (2022, 24-26 June 2022). Support vector machine face recognition application based on particle swarm optimization algorithm. *2022 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA)*. IEEE. https://10.1109/ICAICA54878.2022.9844625

https://doi.org/10.24191/jcrinn.v9i1

Thomas, N., & Kavitha, A. (2020). Prediction of polycystic ovarian syndrome with clinical dataset using a novel hybrid data mining classification technique. *Int J Adv Res Eng Technol (IJARET)*, *11*(11), 1872-1881.

Vedpathak, S., & Thakre, V. (2020). PCOcare: PCOS detection and prediction using machine learning algorithms. *Bioscience Biotechnology Research Communications*, *13*, 240-244. https://doi.org/10.21786/bbrc/13.14/56

Wang, R., Wang, H., Yang, Z., Gui, Y., Yin, Y., & Wang, W. (2021). Recognition of Alzheimer's brain network using hybrid PSO-SVM frame. *2021 40th Chinese Control Conference (CCC)*. IEEE. https://10.23919/CCC52363.2021.9550664

Xue, T., & Jieru, Z. (2022, 15-17 April 2022). Application of support vector machine based on particle swarm optimization in classification and prediction of heart disease. *2022 7th International Conference on Intelligent Computing and Signal Processing (ICSP)*. IEEE. https://10.1109/ICSP54964.2022.9778616