Exploring Machine Learning breakthroughs in PCOS Identification

1st C Hema Prabha

Research Scholar,
Department of Computer Science and Engineering
Dr MGR Educational and Research Institute
Maduravoyal, Chennai - 600095

2nd Dr P S Rajakumar

Professor,
Department of Computer Science and Engineering
Dr MGR Educational and Research Institute
Maduravoyal, Chennai - 600095

3rdDr V Sai shanmugaraja

Professor,
Department of Computer Science and Engineering
Dr MGR Educational and Research Institute
Maduravoyal, Chennai - 600095

1 Abstract

In the field of medicine, Obstetrics and Gynecology play a vital role in humankind. Numerous advancements have been made recently in this subject, such as the application of data science, machine learning, and artificial intelligence to help cure the illnesses that affect women. Machine Learning has played a significant role in predicting Spontaneous Preterm Labor and Birth, predict unplanned cesarean delivery, in gynecologic imaging, in the Diagnosis and classification of polycystic ovarian syndrome, A Predictive Model for the Risk of Infertility in Women, to determine fetal brain ultrasound images as normal or abnormal, gynecological cancer diagnosis to mention a few. The field of Machine Learning is also having exponential growth with new algorithms and models used in Prediction and Diagnosis in every area of healthcare. There has been tremendous growth with new Transformer and GAN models, significantly impacting the field of medicine. This paper aims to perform a systematic review of the role of machine learning and its algorithms in the field of gynecology. A lot of research has happened in recent years and This paper aims to cover the latest developments, emphasizing in identifying the research gap and the evolution of Machine Learning to enhance the clinical experience.

2 Introduction

Polycystic ovary syndrome, or polycystic ovarian syndrome (PCOS), is the most common endocrine disorder in women of reproductive age. It is also one of the pre-eminent causes of infertility, with a worldwide range of 6-26 percent, and in India it is 3.7-22.5 percent. Infertility is defined as the failure in the process of releasing the egg from the ovary. The emergence of an unusual number and volume of follicles during the ovulation phase is one of the several causes of infertility as told by AlAmoudi et al. [1], there are many symptoms, most important including irregular menstruation, excessive hair growth, acne, and obesity. PCOS can be diagnosed through physical examination, such as pelvic ultrasound and blood tests measuring hormone levels. Ayub et al. [2], Diagnosing PCOS is a little tricky as not every person with PCOS has polycystic ovaries (PCO). For this syndrome, experts have developed several criterias. Thus, there isn't a single diagnostic test or method clinicians may use to evaluate individuals. Another factor is that different women experience different symptoms, significantly affecting the Diagnosis. Additionally, symptoms may not always indicate PCOS; instead, they may link to obesity, hypothyroidism, or other endocrine disorders. [3] Risk factors that contribute to the development of PCOS include genetics, neuroendocrine system, sedentary obesity, diet, and way of life.An ultrasound examination of the ovaries should be performed as a result of these symptoms.The specialists recognise the features and morphology of the ovary ultrasound images. It could sometimes be subjective and variable. [1] The ovary contains follicles, which are sac-like cavities. The number of follicles in polycystic ovaries keeps growing.Polycystic ovaries have at least 12 follicles with a 2-9 mm width. The ultrasound image contains a lot of noise, especially Speckle Noise. Because it weakens the image's edge and the finest detectable detail. [4] Long-term effects of diagnosis ambiguity on women's fertility and hormone imbalance may occur. Additionally, the patient may experience inconvenience due to the manual Diagnosis framework. Increased potential for examination errors. Therefore, it is advisable to suggest clever computer- aided systems that can provide gynecologists with tools for decision support.The early Diagnosis of polycystic ovarian syndrome (PCOS) has been made possible by the relatively new technique known as machine learning. Machine learning algorithms can use large patient data sets, including ultra-sound pictures, hormone levels, and demographic information can be used to find trends and forecast a patient's chance of developing PCOS. There are numerous methods (models) for early PCOS detection. The majority of the studies employ SVM, CART, ANN, DT, RF, Gradient Boosting (GB), Gradient Descent (GD), and KNN are the most often used machine learning algorithms. Pushkarini et. al. showed Random Forest algorithm outperforms Linear Regression and KNN by succeeding with lower error val- ues and the highest R^2 value [5]. Daniel et al. examined the effectiveness of several classifiers, includ ing Ensemble Random Forest, Extra Tree, Adaptive Boosting (AdaBoost), and Multi- Layer Perceptron (MLP) uses the dataset with all features and reduced feature sets produced by filter, embedded, and wrapper feature selection techniques. Of these, Ensemble Random For-est achieved 98.98 percent accuracy. [6]Ensemble-based machine learning techniques, such as Random Forest and Extra Trees classifiers have been used to predict PCOS. These methods boost the accuracy and robustness of the model by combining many decision trees.To obtain a final forecast, many decision trees are built and integrated using an ensemble approach known as Random Forest. Each tree in the forest has a random subset of its attributes chosen for it, and the average of the trees' predictions is taken. [2]

It's important to keep in mind that efficient application of machine learning in PCOS detection requires collaboration between researchers, medical professionals, and machine learning specialists.Furthermore, guaranteeing the moral application of data, model interpretability, and validation in diverse populations are critical considerations in developing and deploying machine learning models for PCOS detection.[7]

3 Literature Survey

All papers have used a dataset for PCOS detection. If there is an imbalance in the dataset it can impact the accuracy of the Prediction.SMOTE oversamples the minority samples in the dataset [7] used SMOTE technique to convert 591 samples to 728 samples. For image sample expansion, the data augmentation technique is widely used.[1].Suha et al [8] proposed that the best solution is achieved when the author implemented " VGGNet16", a pre-trained model using CNN archi tecture three as a feature extractor, followed by a stacking ensemble model using the "XGBoost" meta-learner as an image classifier with 99.89% classification accuracy. Lim, J., Li, J., Feng, X., Feng, L. et al. took time-domain parameters, which are commonly used in Traditional Chinese Medicine radial pulse wave analysis. Feng et al. retrieved Thirty-time domain parameters (from both left and right Guan positions) for comparisons: h1, h3, h4, h5, t1, t4, t5, t, w, As, Ad, h3/ h1, h4/h1, h5/h1, w/t. Referring to was used to interpret the meaning of the pulse wave characteristics Lim et. al. [9]Later, these parameters were implemented in different models like KNN,SVM,DT, LR, and LSTM, and the outputs were compared. Zad et al. [10] Statistical Feature selection has helped the authors Zahra Zad et al. improve the AUC score and F1 score in the algorithms they implemented for four different models by different parameters obtained from electronic health records. Naim et al. [11] proposed ten different models for the Prediction of PCOS. It is observed that the hyperparameters selected for each model are different from the other, which has increased the model's performance. Hela et al. proposed a model that used the ensemble model built by combining decisions from several models to improve the model's overall performance. This approach enhances performance over a single model. Zad et al. [10] and Khanna et al. Munjal A. et al.[12] [13]also took the help of a genetic algorithm, a well-known machine learning software called WEKA(Waikato Environment for Knowledge Analysis, developed at the University of Waikato, New Zealand).While most of the papers had used clinical parame ters, Alamoudi et al. Rahman et al.[14] used ultrasound images and implemented the CNN model, the most popular model for Prediction or Detection from images.[15]developed a diagnostic model based on gene expression data using as many samples as possible from the GEO database. Kaur et al. [16] Narinder proposed that Transfer learning has many benefits when classifying PCOS. First, it cuts down on the need for large sets of data. Transfer learning lets you train a model with a much smaller data set than you would need to start from scratch. This is especially helpful for classifying PCOS because the job is so hard that the provided datasets are often small. Some authors have taken ultrasound images to Detect and predict PCOs apart from clinical parameters. Isah Usman et al. [17] implemented segmentation on images and later selected the relevant features using PSO(Particle Swarm Optimization). Pumama et. al. [18], Dewi et. al. [19] used Gabor wavelet, a particular filtering method on segmented images for critical feature extraction. While Pumama et al. [18] implemented three different models, Dewi et. al. [19] used a competitive neural network as the model. Gopalakrishnan et al. [20] has used Scale-Invariant Feature Transform (SIFT), a component discovery calculation in computer vision, to distinguish and portray neighborhoods includes in images. Maheshwari et al. [4]suggest an effective and new F3I approach, which considers three distinct stages: selection for follicle identification, attraction-based ROI selection, and follicle 4 identification. After completing these three stages of feature selection, the feature-selected images are sent into the ANN and NB classifiers, two conventional classifiers. Harshitha et al. [21] aims to develop a Data-driven Computer Aided Diagnostic System (DCADS) using Synthetic Minority Oversampling Technique (SMOTE), PCOS diagnosis using correlation-based feature selection and machine learning approaches without the need for clinical testing. and the essential features are selected above the threshold value of 0.25 using the correlation-based feature selection method. [22] predicted PCOS based only on physical parameters like Anxiety and depression, Social phobia, Body image dissatisfaction, and used fuzzy inference and SVM.

4 Methodology

This section tells about the methodology followed in the review of the papers. There were two review papers that we should have considered for our review, though they formed a basis for our layout of the paper. Around 50 papers dealt with PCOS and implementing ML algorithms. This study involved the datasets used in different papers,the different kind of methodologies implemented by various authors for feature extraction and assessing the algorithms used in their documents. Only considered the recent papers.

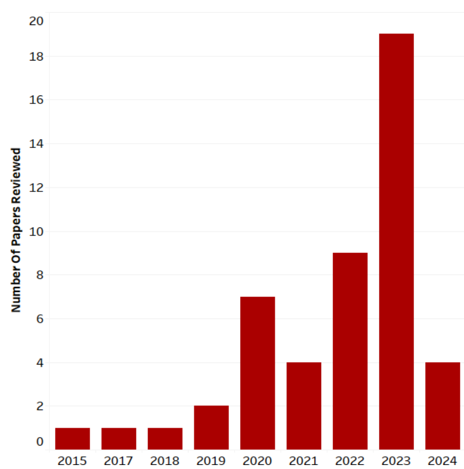

Figure 1: Number of Papers reviewed

The methodology of this paper involved careful selection of the documents. Considered papers that were predicting or diagnosing polycystic ovary syndrome(PCOS) using various machine learning algorithms. Three review papers were not considered for our review. Seven papers could not be validated, or there needed to be quality; hence, we removed those papers from our study.

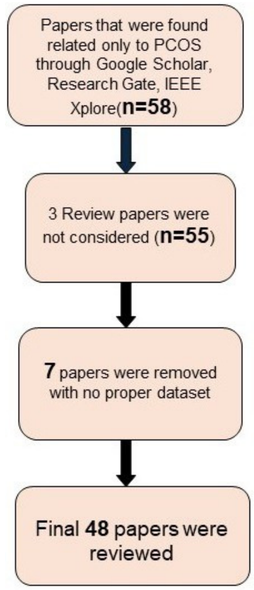

Figure 2: Paper Selection process for this review

Fig.3 illustrates the process of this study. The study involved the analysis of the dataset used in that particular research. Categorized the papers according to the similarity of the dataset.If papers were found using the same dataset, comparing their performance. The study also analyzed the kind of machine learning models implemented and the accuracy of the results. Suitable graphs and tables are tabulated for better understanding. Identified the limitations in each paper . It is observed that there is still a lot of scope for research in the prediction and Diagnosis of PCOS .The areas of research gap has been identified.

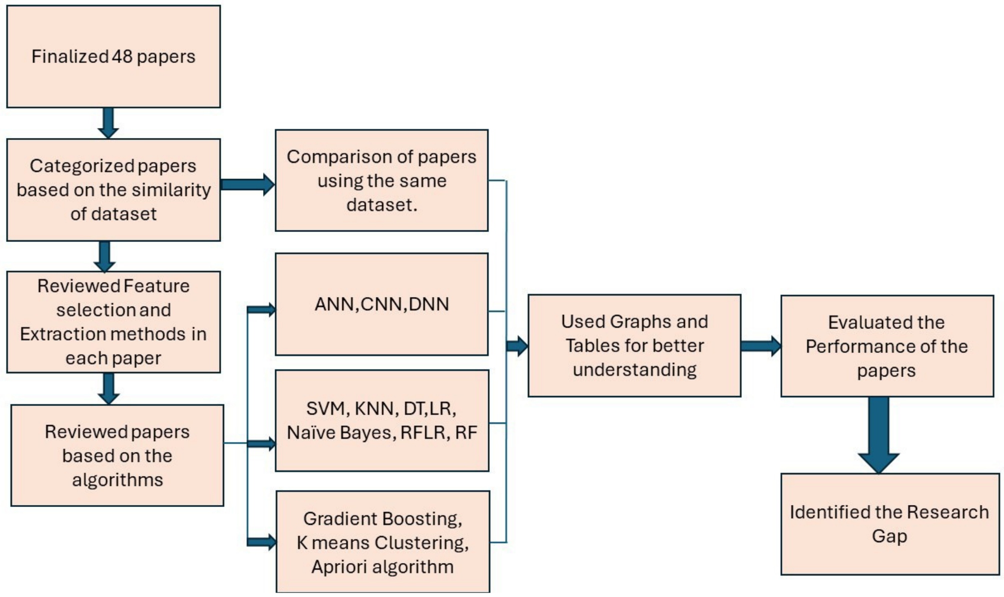

Figure 3: Research Procedure of this work

5 Experiments and Results

5.1 Dataset

The papers studied used several categories of data.Ultrasound images and clinical characteristics have been widely utilised as input for machine learning algorithms.While the study revealed that when the dataset is large ,the model's accuracy is higher.Table 1 gives the details of the datasets .Majority of the datasets size studied is small while [23] had taken a very large dataset of 4700 images and clinical data.When the data is very large ,then big data tools like Kafka and Pyspark can be used. Few papers predicted PCOS only based ultrasound images[18],[19],[20],[24],[8].PCOS diagnosis and Prediction is possible even with only clinical parameters ,for example the Kerala dataset available in kaggle has 41 clinical parameters which is good enough for the Prediction.

Table 1: Details of the Datasets

Datasets Using Clinical Parameters	
A total of 541 samples of patients from ten different hospitals from Kerala, India made the dataset.	[25],[12],[11], [13],[14],[26],[27],[28], [29],[30],[31],[32],[33], [6],[34],[35],[36],[21],[7]
SafetyNet hospital's electronic health records (EHR) from 2003-2016-30,601 women between the ages of 18 and 45,US	[10]
Shangai university of Traditional Chinnese Medicing (SHUTCM) - 459 [2018-22], age - 18-40 (pulse parameters)	[9]
Gene expression data from Gene Expression Omnibus (GEO) database.(264 differentially expressed genes (DEGs))	[15]
Datasets Using Ultrasound Images	
Ultrasound images of Klink Bersalin Permata Bunda Syriah , Cirebon, Indonesia	[18] ,[19]
Nandhini Sri Diagnostic Center,Vellore	[20]
594 images from 2 diagnostic centre and 3 Bangladesh hospitals including 3 combined military hospital(CMH)	[8]
3856 ultrasound images of women aged between 22 and 39 years old, sourced from Kaggle. shared by Anagha Choudhari	[16]
Datasets Using Clinical Parameters and Ultra Sound Images	
1250 patient files,King Fahad Hospital of the University (Khobar, Saudi Arabia),391 images, 127 PCOM, and 264 normal ovaries non-PCOM	[1]
4700 rows of clinical data and ultrasound images of women aged between 15 and 30, sourced from SDM College of Medical Sciences and Hospital, Dharwad. shared by Anagha Choudhari	[23]

5.2 Feature Selection And Extraction

[10]applied a recursive feature elimination approach with L1-penalized logistic regression (L1-regularized RFE) to extract the most informative features and develop parsimonious models.[12] The model performance can be increased if selection of features by employing algorithms like Harris Hawk Optimization, Salp Swarm Algorithm and Mutual Information. [37] used genetic algorithm to extract the most important features from the dataset. [1] Convolution kernel splitting, a 42-layer deep neural network is used by the Inception v3 network to lower parameters and speed up training, making it possible to extract spatial data from smaller convolutions with efficiency.[17]Model performance can be further enhanced by adjusting parameters and by utilizing ensemble techniques, especially Random Forest and Extra Tree Classifier. [10] employed Gradient Boosted Trees (GBT) to determine which elements in decision trees most frequently occurred together. It was discovered that the levels of FSH, LH, SHBG, and estradiol were a significant group of characteristics that are continuous variables and all reproductive hormones that coexisted among trees in all the models. When it comes to ultrasound images, features could be multiple features, including second order features (Contrast, Correlation, Energy and Homogeneity), first order features (Mean, Standard Deviation, Kurtosis, Variance, Skew, and Energy) and geometric features (Diameter, Area, Major Axis Length, Minor Axis Length) were extracted[17].Feature selection done using Particle Swarm Optimization in [17].[18],[19]] used The Gabor wavelet method is applied to produce a filter that is adjustable to the configuration of frequency and orientation utilized. The more combinations used, the more feature vectors produced. However, for easier identification, only the important vectors are in use, i.e., vec- tors that represent PCO and non-PCO follicles. [19]used various segmentation techniques and applied ultrasound image to identify the location of the follicles. Using SIFT algorithm, the feature descriptors were detected.Suha et al[8] used Oriented FAST and Rotated BRIEF(ORB) feature extraction technique which is faster and an effective method.[38]Maheshwari used Furious Flies for Feature Identification(FFFI) comprises three tasks for follicle identification and follicle mapping.[1]A fusion model was used to combine both clinical and image parameters to enhance the feature representation.

Table 2: Feature Selection Techniques Used in Different Studies

Chi Square	[39],[8]
Gabor 2D Filter wavelet method	[18]
SIFT	[19] , [40] ,[34]
Furious Flies for Feature Identification(FFFI)	[38]
Random Forest Embedded Feature Selection	[6]
Heatmap(Correlation Based Selection)	[41],[42]
Kappa values and MSE values	[43]
Select KBEST Algorithm	[44]
PCA	[45],[31],[8]
Univariate Feature Selection Algorithm	[27]
CS-PCOS method based on Goodness of Fit	[11]
Harris Hawks and Salp Swarm Optimization	[12]
Genetic Algorithm	[13]
Mutual Information	[14],[12]
Oriented FAST and Rotated BRIEF(ORB)	[8]

5.3 Algorithm Perfomance

[3] Seven classifiers—KNN, Neural Network, Na¨ıve Bayes, SVM, Classification Tree, Logistic Regression and Linear Discriminant—were implemented in MATLAB. For KNN, using nine nearest neighbors yielded the highest accuracy.

Table 3: The Algorithms Used in Different Studies

Classification Techniques	
Random Forest	[5] , [10], [46], [39], [14], [40] , [47] , [42], [43] ,[44], [45],[34],[15]
Decision Tree	[14], [40]
Gradient Boost	[10] , [40] ,[34]
Support Vectore Machine	[10] ,[46], [41], [39], [14], [18], [40], [42], [44]
Logistic Regression	[5] ,[10], [46], [25] ,[39], [14] ,[40], [47], [42], [44], [17],[34]
KNN	[5], [41] ,[25] ,[18],[34]
Ensemble RF	[6],[39] ,[40], [42] ,[44]
Gaussian Naive Bayes	[41], [39], [42],[34]
XGBoost	[8],[34]
hybrid Logistic regression and random forests (RFLR)	[40]
apriori algorithm	[48]
Neural Networks	
CNN	[12]
DNN	[8]
ANN	[17],[23],[15]
Clustering Techniques	
K means clustering	[40] , [45]

5.4 A Study on the Kaggle Dataset

5.4.1 Dataset and Feature Selection methods used

Around 19 papers in this study has taken the same dataset which is called Polycystic Ovary Syndrome, collected from 10 hospitals across Kerala-India, published 2020, contains clinical and physical parameters of 541 women—364 were in good health, and 177 had a diagnosis of PCOS.This dataset comprised of 44 features Clinical measures included FSH, LH, Hb, TSH, AMH, Vit D3, PRG, and other metabolic, hormonal, and biochemical markers.[6]. While physical parameters included weight, height, age, measurements of waist, hip, and the ratio between the two.[6]In this study three types of feature selection methods, including Sequential Backward Selection (SBS) as a choice of wrapper features method. The PCOS dataset was subjected to the Pearson technique for filter feature selection and the RF method for embed- ded feature selection to assess the impact of feature selection on the performance of different classifers in recognition of patients both with and without PCOS. Feature selection methods reduced the number of features of the original dataset which included 42 features. Pearson filter feature selection method decreased the number of features to 33 and SBS wrapper feature section method reduced the number of features to 30. Additionally, by using RF-embedded features, the PCOS's total feature count was lowered to 28.[13]Khanna et al. used Genetic algo rithm for feature selection to bring down categorical features from 9 to 5 and Numeric features from 31 to 5.Rahaman et al. [14] performed mutual information to select only 12 features.A simple heatmap method was used by Manjunath[41] to remove 17 parameters.[39]Vaidehi used chi-square method for feature selection.[6] devised RF embedded feature selection method and reduced features to 28.[29] implemented a hybrid feature selection approach to reduce the num ber of features using filters and wrappers.[31] used SPSSV22.0 to identify eight potential fea- tures and then transformed with PCA.[27] implemented univariate feature selection algorithm to find the best features in the dataset and found that the most important attributes is the ratio of Follicle Stimulating hormone(FSH) and Leutinizing hormone(LH).Using Kappa values and individual MSE values, 10 features were selected by Moesha Danielle et al. [43].Select KBest Algorithm uses F test scores and p-value to select the most relevant 15 features by [44]Harsh et al.

5.4.2 Comparison of Algorithms Used

The following machine learning models have been applied to this dataset: :

RandomForest : RandomForest : RandomForest: This well-known machine learning method is a member of the supervised learning approach. It can be applied to ML issues involving both classification and regression. It is based on the concept of ensemble learning, which is a process of combining models. Multiple classifiers are used to solve complex problems and improve the performance of the model. Figure 4 shows a comparative analysis of the performance of the random forest algorithm. Random forest has been implemented in most papers.[6] achieved the highest accuracy of 98.89 using the ensemble technique and only 28 features.

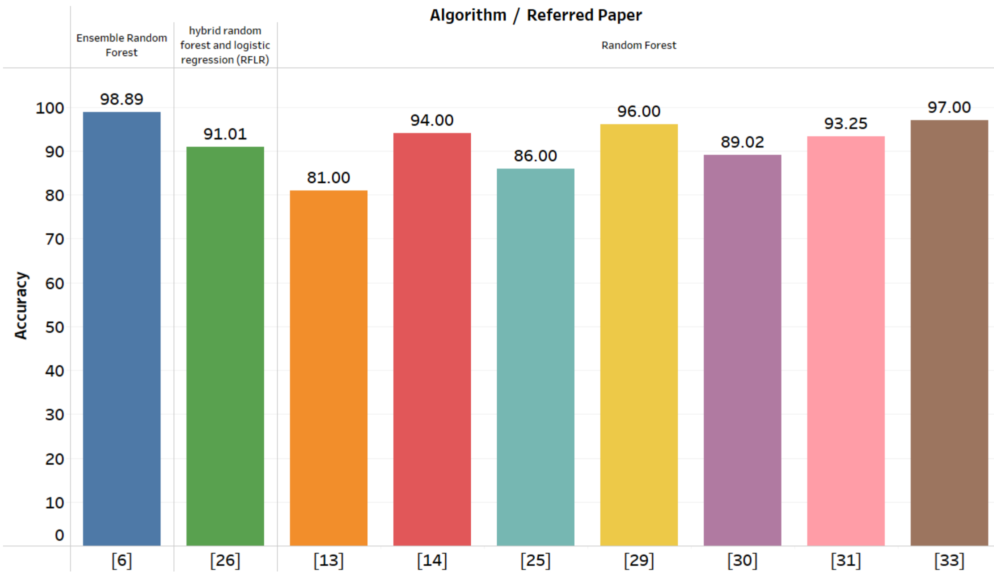

Figure 4: Accuracy Comparison for Kerala Dataset

Naive Bayes : Naive Bayes: Naive Bayes classifiers are a collection of classification algo- rithms based on Bayes' Theorem. It is a family of algorithms rather than a single method, and they are all based on the same principle—that is, each pair of features is not dependent on each other.[11]achieved 100tion approach is proposed based on the optimized chi-squared mechanism. Figure 2 shows the operational flow of the CSPCOS procedure for choosing fea- tures. By comparing the the frequencies that were noticed (categorical data) with the an- ticipated frequency(target data), the suggested CSPCOS technique verifies the independence. The suggested CS-PCOS method retrieves the essential features by evaluating data according to goodness of fit. The suggested feature selection method takes the 39 features as input and calculates the relevance values for each feature. Gaussian naive Bayes is used to demonstrate the feature importance values analysis.

Support Vector Machine(SVM) : SVM algorithms work incredibly well when attempting to determine the largest possible separation hyperplane between the many classes present in the dependent feature.[33] achieved an accuracy score of 93% for SVM with an RBF kernel and multilayer perceptron model.

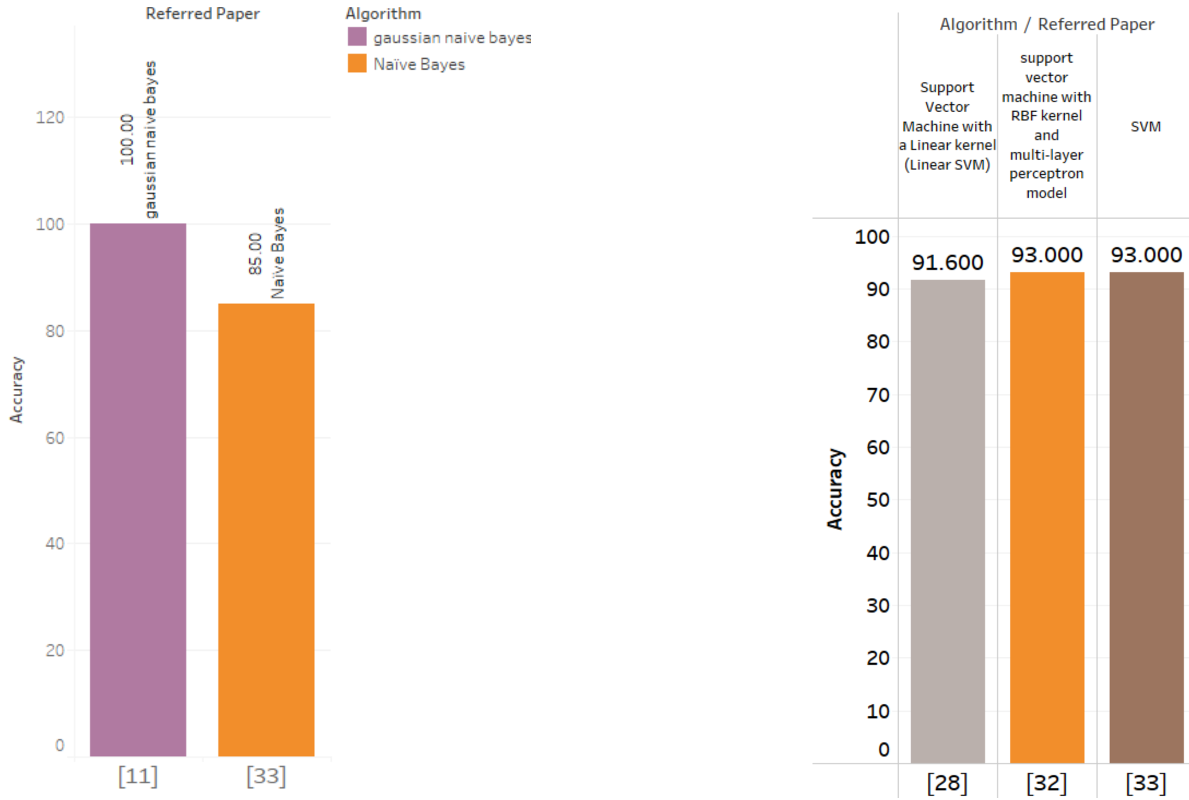

Figure 5: Accuracy Comparison for Kerala Dataset

Algorithms like Extra Trees, Decision Trees, AdaBoost (AB), Logistic Regression, K-means clustering, PCA,KNN, and XGB were also implemented.It is observed that Random forests and Gaussian Naive Bayes models have the highest accuracy in Prediction.

6 Conclusions

From the review comprising of 46 papers, It is evident that there is a vast array of potential areas for research in this domain, and the findings of this review will undoubtedly aid scholars in understanding the depth and breadth of existing studies, requirements and limitations to conduct more experiments and studies for PCOS detection in the future. One major shortcoming is the unavailability of larger dataset which can increase the efficiency of ML models[38],[8],[38].The genetic aspect of the disorder is yet to be unearthed as the major attribute.[13].Features that describe like acne, hirsutism and other signs of hyperandrogenism, amenorrhea and infertility can also be included [1].While PCA has been widely used, Independent Component analysis (ICA) and t-SNE can be explored to further improve feature selection [47].It is also found that pulse wave information has been employed in one among the papers[9].The same can be diversified by integrating with tongue data.Also, ideas related to Explainable AI and federal learning can be incorporated[8].It has been identified that there is still a lot of scope for research in the identification and Diagnosis of Polycystic ovary syndrome (PCOS), using artificial intelligence and machine learning.

References

[1] A. Alamoudi, I. U. Khan, N. Aslam, N. Alqahtani, H. S. Alsaif, O. Al Dandan, M. Al Gadeeb, R. Al Bahrani *et al.*, "A deep learning fusion approach to diagnosis the polycystic ovary syndrome (pcos)," *Applied Computational Intelligence and Soft Computing*, vol. 2023, 2023.

[2] A. Ayub, M. R. Ahmed, N. Kumar *et al.*, "Advancing polycystic ovary syndrome diagnosis: A holistic ensemble-based machine learning approach," in *2023 7th International Conference on Computation System and Information Technology for Sustainable Solutions (CSITSS)*. IEEE, 2023, pp. 1–7.

[3] D. Hdaib, N. Almajali, H. Alquran, W. A. Mustafa, W. Al-Azzawi, and A. Alkhayyat, "Detection of polycystic ovary syndrome (pcos) using machine learning algorithms," in *2022 5th international conference on engineering technology and its applications (IICETA)*. IEEE, 2022, pp. 532–536.

[4] K. Maheswari, T. Baranidharan, S. Karthik, and T. Sumathi, "Modelling of f3i based feature selection approach for pcos classification and prediction," *Journal of Ambient In- telligence and Humanized Computing*, vol. 12, pp. 1349–1362, 2021.

[5] H. Pushkarini and M. Anusuya, "A prediction model for evaluating the risk of developing pcos," *Int Res J Eng Technol (IRJET)*, vol. 7, no. 9, pp. 1150–6, 2020.

[6] H. Danaei Mehr and H. Polat, "Diagnosis of polycystic ovary syndrome through different machine learning and feature selection techniques," *Health and Technology*, vol. 12, no. 1, pp. 137–150, 2022.

[7] E. Nsugbe, "An artificial intelligence-based decision support system for early diagnosis of polycystic ovaries syndrome," *Healthcare Analytics*, vol. 3, p. 100164, 2023.

[8] S. A. Suha and M. N. Islam, "An extended machine learning technique for polycystic ovary syndrome detection using ovary ultrasound image," *Scientific Reports*, vol. 12, no. 1, p. 17123, 2022.

[9] J. Lim, J. Li, X. Feng, L. Feng, Y. Xia, X. Xiao, Y. Wang, and Z. Xu, "Machine learning classification of polycystic ovary syndrome based on radial pulse wave analysis," *BMC Complementary Medicine and Therapies*, vol. 23, no. 1, p. 409, 2023.

[10] Z. Zad, V. S. Jiang, A. T. Wolf, T. Wang, J. J. Cheng, I. C. Paschalidis, and S. Mahalingaiah, "Predicting polycystic ovary syndrome (pcos) with machine learning algorithms from electronic health records," *medRxiv*, pp. 2023–07, 2023.

[11] S. Nasim, M. S. Almutairi, K. Munir, A. Raza, and F. Younas, "A novel approach for polycystic ovary syndrome prediction using machine learning in bioinformatics," *IEEE Access*, vol. 10, pp. 97 610–97 624, 2022.

[12] V. V. Khanna, K. Chadaga, N. Sampathila, S. Prabhu, V. Bhandage, and G. K. Hegde, "A distinctive explainable machine learning framework for detection of polycystic ovary syndrome," *Applied System Innovation*, vol. 6, no. 2, p. 32, 2023.

[13] A. Munjal, R. Khandia, and 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," *Int J Sci Res*, vol. 9, pp. 1–5, 2020.

[14] M. M. Rahman, A. Islam, F. Islam, M. Zaman, M. R. Islam, M. S. A. Sakib, and H. M. H. Babu, "Empowering early detection: A web-based machine learning approach for pcos prediction," *Informatics in Medicine Unlocked*, p. 101500, 2024.

[15] N.-N. Xie, F.-F. Wang, J. Zhou, C. Liu, and F. Qu, "Establishment and analysis of a combined diagnostic model of polycystic ovary syndrome with random forest and artificial neural network," *BioMed research international*, vol. 2020, no. 1, p. 2613091, 2020.

[16] N. Kaur, G. Gupta, and P. Kaur, "Transfer-based deep learning technique for pcos detection using ultrasound images," in *2023 International Conference on Network, Multimedia and Information Technology (NMITCON)*. IEEE, 2023, pp. 1–6.

[17] O. R. Isah, A. Usman, and A. Tekanyi, "A hybrid model of pso algorithm and artificial neural network for automatic follicle classification," 2017.

[18] B. Purnama, U. N. Wisesti, F. Nhita, A. Gayatri, T. Mutiah *et al.*, "A classification of polycystic ovary syndrome based on follicle detection of ultrasound images," in *2015 3rd International conference on information and communication technology (ICoICT)*. IEEE, 2015, pp. 396–401.

[19] R. Dewi, Adiwijaya, U. N. Wisesty, and Jondri, “Classification of polycystic ovary based on ultrasound images using competitive neural network,” in *Journal of Physics: Conference Series*, vol. 971. IOP Publishing, 2018, p. 012005.

[20] C. Gopalakrishnan and M. Iyapparaja, “Detection of polycystic ovary syndrome from ultrasound images using sift descriptors,” *Bonfring International Journal of Software Engineering and Soft Computing*, vol. 9, no. 2, pp. 26–30, 2019.

[21] H. Batra and L. Nelson, “Dcads: Data-driven computer aided diagnostic system using machine learning techniques for polycystic ovary syndrome,” *International Journal of Performability Engineering*, vol. 19, no. 3, p. 193, 2023.

[22] A. Kodipalli and S. Devi, “Prediction of pcos and mental health using fuzzy inference and svm,” *Frontiers in public health*, vol. 9, p. 789569, 2021.

[23] A. Kodipalli, S. Devi, and S. Dasar, “Semantic segmentation and classification of polycystic ovarian disease using attention unet, pyspark, and ensemble learning model,” *Expert Systems*, vol. 41, no. 3, p. e13498, 2024.

[24] N. Kaur, G. Gupta, and P. Kaur, “Transfer-based deep learning technique for pcos detection using ultrasound images,” in *2023 International Conference on Network, Multimedia and Information Technology (NMITCON)*. IEEE, 2023, pp. 1–6.

[25] N. Tanwani, “Detecting pcos using machine learning,” *Int J Modern Trends Eng Sci (IJMTES)*, vol. 7, no. 1, pp. 1–20, 2020.

[26] I. Silva, C. Ferreira, L. Costa, M. Sóter, L. Carvalho, J. de C. Albuquerque, M. Sales, A. Candido, F. Reis, A. Veloso *et al.*, “Polycystic ovary syndrome: clinical and labo- ratory variables related to new phenotypes using machine-learning models,” *Journal of Endocrinological Investigation*, pp. 1–9, 2022.

[27] S. Bharati, P. Podder, and M. R. H. Mondal, “Diagnosis of polycystic ovary syndrome using machine learning algorithms,” in *2020 IEEE region 10 symposium (TENSYMP)*. IEEE, 2020, pp. 1486–1489.

[28] A. Zigarelli, Z. Jia, and H. Lee, “Machine-aided self-diagnostic prediction models for polycystic ovary syndrome: Observational study,” *JMIR Form Res*, vol. 6, no. 3, p. e29967, Mar 2022. [Online]. Available: https://formative.jmir.org/2022/3/e29967

[29] Y. A. A. Adla, D. G. Raydan, M.-Z. J. Charaf, R. A. Saad, J. Nasreddine, and M. O. Diab, "Automated detection of polycystic ovary syndrome using machine learning techniques," in *2021 Sixth international conference on advances in biomedical engineering (ICABME)*. IEEE, 2021, pp. 208–212.

[30] M. M. Hassan and T. Mirza, "Comparative analysis of machine learning algorithms in diagnosis of polycystic ovarian syndrome," *Int. J. Comput. Appl*, vol. 975, p. 8887, 2020.

[31] A. Denny, A. Raj, A. Ashok, C. M. Ram, and R. George, "i-hope: Detection and prediction system for polycystic ovary syndrome (pcos) using machine learning techniques," in *TENCON 2019-2019 IEEE Region 10 Conference (TENCON)*. IEEE, 2019, pp. 673–678.

[32] S. Tiwari, L. Kane, D. Koundal, A. Jain, A. Alhudhaif, K. Polat, A. Zaguia, F. Alenezi, and S. A. Althubiti, "Sposds: A smart polycystic ovary syndrome diagnostic system using machine learning," *Expert Systems with Applications*, vol. 203, p. 117592, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0957417422009046

[33] P. Bhardwaj and P. Tiwari, "Manoeuvre of machine learning algorithms in healthcare sector with application to polycystic ovarian syndrome diagnosis," in *Proceedings of Academia-Industry Consortium for Data Science*, G. Gupta, L. Wang, A. Yadav, P. Rana, and Z. Wang, Eds. Singapore: Springer Nature Singapore, 2022, pp. 71–84.

[34] B. Yamini, V. R. Kaneti, M. Nalini, and S. Subramanian, "Machine learning-driven pcos prediction for early detection and tailored interventions," *SSRG International Journal of Electrical and Electronics Engineering*, vol. 10, pp. 61–75, 2023.

[35] L. Sanjay, N. Nayak, D. Sriram, A. Kodipalli, T. Rao, and S. Kamal, "Analysis and prediction of polycystic ovarian syndrome using ml classifiers check for updates," *Advances in Communication and Applications: Proceedings of ERCICA 2023, Volume 2*, vol. 1105, p. 261, 2024.

[36] D. Hdaib, N. Almajali, H. Alquran, W. A. Mustafa, W. Al-Azzawi, and A. Alkhayyat, "Detection of polycystic ovary syndrome (pcos) using machine learning algorithms," in *2022 5th International Conference on Engineering Technology and its Applications (IICETA)*, 2022, pp. 532–536.

[37] F. Faridoon, R. H. Ali, Z. Ul Abideen, N. Shahzadi, A. Z. Ijaz, U. Arshad, N. Ali, M. Imad, and S. Nabi, "Prediction of polycystic ovary syndrome using genetic algorithm-driven feature selection," in *2023 International Conference on IT and Industrial Technologies (ICIT)*, 2023, pp. 1–6.

[38] K. Maheswari, T. Baranidharan, S. Karthik, and T. Sumathi, "Modelling of f3i based feature selection approach for pcos classification and prediction," *Journal of Ambient In- telligence and Humanized Computing*, vol. 12, pp. 1349–1362, 2021.

[39] V. Thakre, S. Vedpathak, K. Thakre, and S. Sonawani, "Pcocare: Pcos detection and prediction using machine learning algorithms," *Biosci Biotechnol Res Commun*, vol. 13, no. 14, pp. 240–244, 2020.

[40] S. Aggarwal and K. Pandey, "Early identification of pcos with commonly known diseases: Obesity, diabetes, high blood pressure and heart disease using machine learning techniques," *Expert Systems with Applications*, vol. 217, p. 119532, 2023.

[41] M. Alagarsamy, N. Shanmugam, D. P. Mani, M. Thayumanavan, K. K. Sundari, and K. Suriyan, "Detection of polycystic syndrome in ovary using machine learning algorithm," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 11, no. 1, pp. 246–253, 2023.

[42] S. Kaushik and S. K. Mishra, "Comparative analysis of classification methods for pcod prediction," in *2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT)*. IEEE, 2023, pp. 1–8.

[43] M. D. G. Moore, K. Baboolal, and P. Hosein, "On the prediction of polycystic ovarian syndrome," in *2023 IEEE International Conference on Technology Management, Operations and Decisions (ICTMOD)*. IEEE, 2023, pp. 1–6.

[44] H. Pathak, P. Handa, and N. Goel, "Comparative analysis of feature selection methods for automatic classification of pcod," in *2023 10th IEEE Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON)*, vol. 10, 2023, pp. 235–241.

[45] P. Karadayı Ata¸s, "Exploring the molecular interaction of pcos and endometrial carcinoma through novel hyperparameter-optimized ensemble clustering approaches," *Mathematics*, vol. 12, no. 2, p. 295, 2024.

[46] H. Batra and L. Nelson, "Dcads: Data-driven computer aided diagnostic system using machine learning techniques for polycystic ovary syndrome," *International Journal of Performability Engineering*, vol. 19, no. 3, p. 193, 2023.

[47] D. Kumar and A. Kumar, "Pcos prediction using machine learning techniques," *NEU Journal for Artificial Intelligence and Internet of Things*, vol. 1, no. 2, 2023.

[48] S. Pradeepa, K. Geetha, K. Kannan, and K. Manjula, "Deodorant: a novel approach for early detection and prevention of polycystic ovary syndrome using association rule in hypergraph with the dominating set property," *Journal of Ambient Intelligence and Humanized Computing*, pp. 1–17, 2023.