Multimedia Tools and Applications
https://doi.org/10.1007/s11042-024-19707-6

Polycystic Ovary Syndrome (PCOS) diagnostic methods in machine learning: a systematic literature review

Shikha Arora[1] · Vedpal[2] · Naresh Chauhan[3]

Received: 29 August 2023 / Revised: 3 June 2024 / Accepted: 15 June 2024

Abstract
Polycystic Ovarian Syndrome, also known as PCOS, is a major hormonal imbalance affecting women primarily in their reproductive age. Women with PCOS may have either infrequent or extended menstrual cycles or sometimes excess male hormone i.e. androgen levels. The ovaries may grow with number of slight collections of fluid, called follicles that fail to release eggs every month regularly. It is also seen that PCOS not only affects a woman's fertility but also indirectly causes various other health issues like type 2 Diabetes, obesity, blood pressure and other metabolic disorders. Recent researches have focussed on the use of different algorithms in Machine Learning to diagnose PCOS using structured or unstructured data. Therefore, in this study, a considerable Literature Review has been carried out to provide a detailed analysis of various algorithms that have been used to detect PCOS with their comparative study. Also some of the work particularly focuses on health issues, the food and dietary patterns and the ways to manage PCOS. Further, these algorithms are critically analysed to understand the framework and limitations that should be considered while putting forward the solutions relevant to the diagnosis of PCOS in an effective way.

Keywords Machine learning · Polycystic ovary · Data science · Deep learning · Uterine fibroids · Follicles · Classification · Regression · SVM · Random forest · CNN · AdaBoost

1 Introduction

The term, Polycystic Ovary Syndrome (PCOS), was first defined by the scientists [1] Stein and Leventhal as an endocrine disorder and since then, it has been identified as one of the major causes of infertility in women. It is a hyperandrogenic disorder related with

✉ Vedpal
ved_ymca@yahoo.co.in

1 J.C. Bose University of Science and Technology, YMCA, Faridabad, Haryana, India

2 Department of Computer Applications, J.C. Bose University of Science and Technology, YMCA, Faridabad, Haryana, India

3 Department of Computer Engineering, J.C. Bose University of Science and Technology, YMCA, Faridabad, Haryana, India

Published online: 27 June 2024

persistent [1] oligo-anovulation and polycystic physiology of the ovaries. It is further correlated with psychological disablement, such as depression and other mood related disarray, metabolic disorganisation, mainly insulin resistance, identified as a major element responsible for modified display of androgen hormone levels and metabolism.

It is a hormonal imbalance which is common mainly in women of reproductive age till their later stages in life. Women with PCOS may suffer from infrequent or extended menstrual periods or excess male hormone or androgen levels. The ovaries may grow with number of small collections of fluid filled follicles and fail to release eggs on a regular basis. PCOS is a "syndrome," or collection of symptoms grouped together that affect the ovaries and ultimately ovulation. The three main parameters related with it are:

- Infrequent, extended or skipped menstrual cycle.
- Multiple cysts in the ovaries
- High levels of androgens which are the male hormones.

Artificial Intelligence is a branch of Computer Science that combines various aspects like learning, reasoning, Knowledge representation, problem solving etc. It helps in creating machines that can act like humans with their ability to reason and solve problems. This important feature of AI has led to the evolution of its sub-field called Machine Learning. Machine Learning has a huge potential in the Healthcare sector as it can work with large amounts of datasets – structured or unstructured, recognise minute details from the images and predict accurate outcomes from them. It has been used by many people and they have succeeded in implementing it when the right amount of dataset is fed to the algorithm.

The description of PCOS has resulted in a splendid rise of medical and technical interest in this disorder, which should be enhanced later on to improve respective medical perspectives and, ultimately devise remedial strategies. To further discuss and interchange views on PCOS, an international group of researchers in this field, assemble every other year to discuss the situation and focus on further research.

PCOS is a binary classification problem which means its clear focus is on determining whether a woman has acquired the symptoms that cause PCOS or not. There are no other choices. Also medical practitioners examine it by counting the number of follicles measured on transvaginal ultrasound images of ovary. This is a time-consuming process and may lead to human errors sometimes.

Machine Learning is a promising approach for PCOS detection as the existing methodologies and treatments are insufficient. Its early diagnosis by the use of ML models can not only save time for clinicians and other healthcare providers but also substantially reduce further disease problems and long-term complications. Most of the reviews like [33], and [98] in this paper focus on such articles, which strongly emphasize the use of Machine Learning models for classifying PCOS. Different researches have also shown that people showing signs of PCOS, also show great evidence of other diseases like diabetes, metabolic inflammation, kidney and heart disorders etc. With the help of Machine Learning, the different algorithms can not only help in predicting PCOS but also other such related disorders. Thus, their early prediction can help in creating Machine Learning models which further aid healthcare providers in suggesting effective diet plans for such diseases. Bulsara et al. [18] clearly states that bringing some lifestyle changes can improve the condition effectively.

In this review, we try to highlight the various methods used by researchers in this field of medicine. Some have used Dimensionality reduction techniques, while others have used supervised learning, deep learning or neural networks for classifying this problem.

1.1 Motivation

After going through many areas of Medicine where Machine Learning can be applied, work toward the problem of PCOS diagnosis is considered for the study, as this is a problem affecting women of all ages. With the help of Machine Learning, reliable models can be selected that can not only help in early diagnosis of the problem but can prevent long term health complications. Some of the key reasons that led to this motive are:

a) This is one of the leading issues that are affecting women of reproductive age in the present time.
b) If the problem is diagnosed well in time, using a cost effective method, it can benefit the clinicians and the patients. There are various parameters in Machine Learning like Precision, Recall, Accuracy etc. that can verify the effectiveness of the model to make it more prominent in classification.
c) Early stage awareness can help patients to take corrective actions well in time. This can prevent any further complications of the disease from occurring.
d) The various papers presented in this review emphasized on the seriousness of this issue as it does not only affects women in conceiving but also leads to other metabolic and heart problems. Machine Learning has proved to be effective in not only managing the symptoms of PCOS but also reduce other health related risks.
e) This problem leads to higher stress and anxiety levels as well. An early detection and prediction model has been proposed in reference [36] that focuses on developing Machine Learning model that not only detects PCOS but also relates it to the psychological well being of a woman. Thus there is a strong need for improved protocols and guidelines developed through ML for balancing stress and mental health of a woman.

1.2 Our major contributions

The majority part of this paper is focused towards:

- An extensive analysis of different categories of Machine Learning Algorithms used in identifying PCOS problems.
- Detailed inspection of various researches proposed by the researchers that helped in identifying PCOS effectively.
- Examine the major features or parameters used by the researchers that have helped them to make their study effective.
- A comparative study on the focus areas given by various researchers, that could help in taking into account the research gap and future considerations.
- Summarise the effect of PCOS on a woman's health – directly or indirectly.

1.3 Paper structure

The research paper is further arranged as follows:

Section 2 provides an outline of how Machine learning advanced with time. Section 3 presents the related research existing in the mentioned literature. Further in Section 4 the detailed research strategy to be adopted for performing the Systematic Literature Review is described. It includes a complete approach for the stated topic. Also it tries to focus on the major findings and the comparative study between the different Machine Learning

Approaches used for PCOS analysis is given. Then we have discussed the major gap areas that still continue to exist, leading to some more research directions in the future. Finally, in Section 5 at the end of the paper, a clear conclusion is drawn by clearly drawing attention towards the future scope.

2 Machine learning evolution

The Fig. 1 gives a brief description of how Machine Learning evolved with time. Although it is a branch of Artificial Intelligence but it is in itself a vast area to be studied. So it is important to analyse how Machine Learning has been adopted by people in the coming few years at such a faster rate.

Machine Learning has evolved of its own as a major subdivision of Artificial Intelligence(AI). It can help computer systems in achieving better accuracy and performance over the various tasks related to decision making, classification etc. It has become an efficient response tool in fields like e-commerce and cloud computing.

Alan Turing in the year 1936, created a very powerful hypothetical machine called Turing Machine. It could help in solving computational problems through an infinite tape and also check for problems that did not have any solution. Around 1951–1952, Marvin Lee Minsky developed Stochastic Neural Analog Reinforcement Calculator (SNARC) that was the first ever artificial neural network computer which could work through trial and error, learn from experience, thus improving its performance. Further in 1950s Arthur Samuel created a wonderful board game called Checkers between computer and player. Some random moves were chosen and based on learning and training the computer improved its performance.

Perceptron, a single layer Artificial Neural was then invented by Frank Rosenblatt. It could work on the principle of calculating the output based on the inputs and weights and comparing them with the threshold value. Then Machine Learning further evolved by developing supervised algorithms that could help in classifying the input data. These

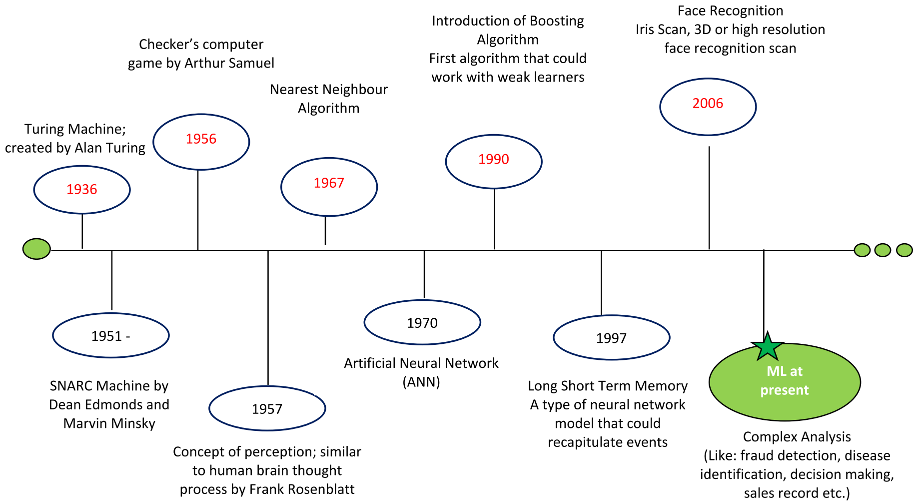

Fig. 1 Machine learning evolution

algorithms are like Logistic Regression, K- Nearest Neighbour, and Decision Tree etc. Artificial Neural Network works on the concept of Biological Neuron but it is a multilayer perceptron and takes decisions collectively. It was introduced in the year 1970.

In 1990, the concept of Ensemble Learning was introduced that worked on the principle of bagging and boosting. Here the output of various learners was combined to form a strong learner that could learn from its surroundings and help in the process of classification. In 1997 Long Short Term Memory Neural Network was created that could work by recapitulating events.

In the year 2006, the Machine Learning Algorithms started working on face recognition, Iris Scan, 3D scan etc. and generated such powerful devices that could easily implement these features. In the present era, ML is highly recommended in areas like disease prediction, sales record management, decision making, retail and marketing, business analysis and the like.

2.1 Introduction to machine learning

Machine Learning [2] is a subfield of Artificial Intelligence, which is defined as the capability of machine to imitate intelligent human behaviour. It allows various software applications to make accurate predictions. In the present world, huge amount of digital data is produced, especially the health data. To analyse this data, and to develop smart algorithms that can help in correct diagnosis, Artificial Intelligence and thus Machine Learning is the key. Various types of Machine Learning Algorithms exist, which present a great impact on the type of health data being analysed.

2.1.1 Supervised machine learning

It is a type of Machine Learning that uses labelled training data and a collection of training examples to infer the result that maps an input to an output based on given input – output pairs. The algorithms that are covered under supervised learning approach are Classification and Regression.

Classification algorithms are helpful when the output variable is a categorical value – like presence or absence of disease, etc. The various classification models are Logistic Regression, Naive Bayes, and Random Forest etc.

Logistic Regression uses a Sigmoid function given by the equation

$$Y = 1/\left(1 + e^{-(a0+a1x)}\right) \tag{1}$$

where, Y is the predicting variable, a0 is the intercept and a1 is the coefficient which shows the effect of incrementing or decrementing x on Y, x is the independent variable.

Naïve Bayes Theorem works on the principle of Bayes theorem.

$$\text{It is expressed mathematically as } P(Y/X1, X2, X3 \ldots .Xn) = \left(P(X1|Y)^{*}P(X2|Y)^{*}P(X3|Y)^{*} \ldots \ldots {}^{*}P(Xn|Y)^{*}P(Y)\right) / \left(P(X1)^{*}\, P(X2)^{*}P(X3) \ldots ..^{*}P(Xn)\right) \tag{2}$$

The Eq. (2) expresses the probability of existence of a dependent variable Y, which depends on multiple independent variables X1,X2,X3…Xn.

Similarly we evaluate the probability of non-existence of the variable Y. Thus the algorithm helps us in classification.

Random Forest is an ensemble learning method (where group of Decision Trees are considered to evaluate the result) for Classification. The results of these decision trees classification is then decided using the concept of majority voting.

Regression algorithms are used when the output variable is a real or continuous value. It explains how a variable (dependent) changes as the other variable (independent) changes. The various regression models are linear regression and Multiple Regression.

$$\text{Linear Regression is mathematically represented as } Y = mX + b \tag{3}$$

The Eq. (3) shows that.

Y is the dependent variable;
X is the independent variable;
m is the slope of the line, that shows how much Y changes for a unit change in X;
b is the intercept which depicts the value of Y when X is 0.

2.1.2 Unsupervised machine learning

It is the training of a machine using information that is neither classified nor labelled and allowing the algorithm to act on that information without guidance. Here the task of the machine is to group unsorted information according to similarities, patterns, and differences without any prior training of data. It includes techniques like Clustering and Association.

Clustering tries to find common features among the data objects and based on the similarities it groups them into clusters.

Association Rules are basically used in Market Basket Analysis which finds relationship between variables i.e. the set of data items that occur together in the dataset.

2.1.3 Reinforcement machine learning

Reinforcement is the act of encouraging or pushing someone to attain a particular task. It is an area of Machine Learning where there is an agent, that behaves in a particular way (called action), to maximize its reward. Recently it has been used in many health care applications to check symptoms in disease diagnosis.

2.1.4 Deep learning

It is that [3] subset of machine learning, which is essentially based on the structure of neural network with three or more than three layers. These neural networks strive to imitate the behaviour of the human brain thus enabling it to “learn” from a larger amount of data. While a neural network with a single layer can make predictions that are approximate values, additional hidden layers can help in optimizing and refining the procedure for higher accuracy. Deep learning is distinguished from classical machine learning based on the type of data that it works with and the ways in which it learns. Deep learning also helps in eliminating some of the data pre-processing steps which are typically involved in machine learning. These algorithms can consume and process unstructured data, like text and images, and they also automate feature extraction, removing most of the dependencies on technical experts.

Neural networks Neural networks, the term which is also known as Artificial Neural Networks (ANNs) or Simulated Neural Networks (SNNs), are that subset of machine learning which forms the heart of deep learning algorithms [4]. Their name and structure are inspired by the human brain, mimicking the way the biological neurons signal to one another. Neural networks rely on the given training data to learn and improve its accuracy with time. However, once these algorithms are calibrated for accuracy, they prove to be very powerful tools in computer science and machine learning allowing us to classify and cluster data with a higher velocity. Some common tasks like speech recognition or recognising images can take minutes to several hours when compared to the manual identification by technical experts. One of the most famous neural networks ever known is Google's search algorithm. Garg et al. [5] also presented a survey on the role of machine learning in research reacted to the medical. The Fig. 2 demonstrates briefly the classification of Machine Learning Algorithms and their further subdivisions.

3 Related surveys

In this section rewrite the overview of the various proposed methods to diagnose polycystic ovary syndrome (PCOS) using Machine Learning.

The work done in [6, 7, 9] deal with traditional methods and ensemble classifiers like SVM, Random Forest, ADA Boost, KNN, Naïve Bayes, Decision Tree etc. to diagnose the presence of polycystic ovaries. Each of these classifiers was then evaluated for accuracy after putting in certain feature selection methods like wrapper, filter and embedded methods. The study could be further extended by including the use of different types of data sets for diagnosing PCOS. It is also suggested that a web portal can be created where, the patients, unaware of the symptoms, can get themselves checked, and after getting tested clinically, can get their PCOS status validated.

Ahmed et al. [8] have used VGG016 for classification and followed it with comparative analysis of previous technologies. They have used TCGA-UCEC dataset for Uterine Fibroids Classification. Image Segmentation was done to produce the final image. Results

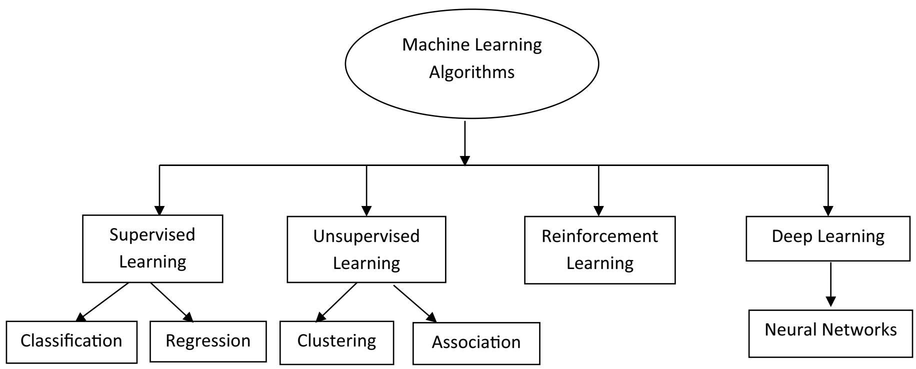

Fig. 2 Types of machine learning algorithms

showed that the VGG 16 is 98.5% confident to detect the presence of uterine fibroids from images with 0.2% loss and is also the best deep learning model when compared with previous accuracies.

Zigarelli et al. [10] have developed two models, namely—the patient model (based on nonsurgical measures such as anthropomorphic measures, age, symptoms, age, and other lifestyle factors) and provider model(based on their medical test results and all other non-invasive measures). Both models were compared on the basis of various error metrics. The dataset for this study was collected from 10 different hospitals in Kerala. They adopted the CatBoost method for performing classification, K-fold cross-validation was used for estimating the performance of the models, and SHAP (Shapley Additive Explanations) values were used to explain the importance of each variable. They also used k-means clustering and Principal Component Analysis (PCA) to split the data set into 2 distinct BMI subgroups and compared the results predicted as well as the feature importance between the 2 subgroups. It resulted in prediction accuracy about the PCOS status as 81% to 82.5% without using any invasive measures in the patient models and 87.5% to 90.1% prediction accuracy when both noninvasive and invasive predictor variables were used in the provider models.

Nautiyal et al. [11] explained how the genes directly or indirectly affect the growth and progression of PCOS. They have also identified four types of phenotypic classification of PCOS which further categorizes PCOS patients as Classic, Ovulatory and non-Hyperandrogenic patients. Due to the complexity of disease and variation on symptoms, there is no specific treatment and personalized attention is required by every patient. They have recommended that early diagnosis is possible with genetic variance that can help in finding a better treatment.

Sarvamangla et al. [12, 25] and Bhosale et al. [13] have carried out a study on the use of CNN and DCNN in medical imaging and diagnosis process like segmentation, localization, classification, and detection. Also, the application of CNN in ailments of various organs like the brain, breast, lungs etc. has been critically analyzed. In future work, the results of CNN can be combined with a prediction system that could verify the decisions taken. Image captioning can also be included to help and guide physicians to understand the network in a better way. Yang et al. [16] focused on object detection and semantic segmentation as the two main branches of the traditional CNN model in image analysis. In medical science, various types of images are produced, like radiation images (CT scan, MRI etc.), pathological images, ultrasound images, endoscopic images etc. Different CNN models like VGG16, GoogleNet etc. can be selected for feature extraction. The limitation of these techniques is that a large number of images is needed for training them and also the help of a doctor is needed to label the images. Future studies can focus on the use of limited medical images with good training results. Sumathi et al. [31] have used CNN as an image classifier. They have used ultrasound images collected independently through some sources. Further, the test data is used in feature extraction and segmentation to classify cysts. A hybrid model has been proposed by Suha et al. [98] that combined the benefits of CNN with transfer learning as a method for extracting features and stacked ensemble machine learning classifiers for detecting PCOS from ultrasound images. The proposed technique is also compared with other existing methodologies for evaluating its efficiency. The study will benefit healthcare providers in quick and accurate identification of PCOS. The low point of this study is that the hybrid model generated with the fusion of deep learning and stacked ensemble Machine Learning classifier is indicated as something unexplained for the physicians. In future, the same work can be performed using Federated Learning and some Explainable AI (XAI).

Rajeev et al. [14] aimed at providing a comparative study of psychological implications like stress, anxiety, depression etc. on 65 women suffering from PCOS and 59 without PCOS. It was concluded that women suffering with PCOS have higher chances of acquiring such mental issues. Future work could focus on identifying those aspects or symptoms of PCOS that contribute most to psychological factors. Also a study on relation between biological issues and anxiety and stress factors could be analysed. Gopalakrishnan et al. [15] aimed to propose an effective PCOS detection and classification system. The system consisted of certain processes like an ultrasound image pre-processing, follicle segmentation and classification of ovarain type. Using guassian low pass filter, the image was initially processed and with the help of multilevel thresholding technique, the follicles were identified in it. The important features from the image were then extracted with geometric properties of the detected follicles, the feature vector was used to train the supervised machine learning classifiers. The results showed that better accuracy was achieved with SVM classifier using RBF kernel.

Isaza et al. [17] have presented an overview of certain deep learning methods assisted by the mathematical models that explain their functionality. They also demonstrate some key elements that can be used in modern image classification networks. They have even shown some applications recognized in the past years. The development of such networks is limited by the type of networks they are trained on. As a future scope, they propose efficient network or optimization methods that can work on less training data, perform data augmentation to increase the availability of data or carry out image synthesis to improve network performance that has operated on very few reference images. Bulsara et al. [18] in their review extended that PCOS is a composite condition indicated by anovulation, cysts in the ovaries, and endocrine variation. It emphasizes more on the clinical symptoms, risk factors involved and the pathophysiological factors involved rather than finding a cure for it. The authors clearly state that bringing some lifestyle changes can ease the symptoms related with PCOS. Qayuum et al. [19] have provided an overview of various security and privacy related challenges faced in healthcare sector. Also some methods like Homomorphic Encryption, multi- party computation, differential privacy etc. are presented that depict secure use of ML in such medical applications. They have also mentioned some countermeasures against adversarial attacks and grouped them into 3 classes – modifying the model, modifying the data and introducing an auxiliary model(s). At the end, some challenges observed in the current research and future directions were given for future research.

Chen et al. [20] have elaborated on the various ethical concerns with the use of machine learning. They also propose a pipeline of ethical machine learning (ML) in health that highlights both visible and hidden challenges. The pipeline clearly states the disparities and biasing in its various stages like selecting Problem, collecting Data, defining Outcome, developing Algorithm and Post deployment considerations. They have also brought up some issues to be addressed in future. This includes how to encourage ML model developers to include ethical considerations into the streamline from the initial stages, how to convert these evaluations of ML systems into generating meaningful practices that could be used clinically, how to use ML specifically for the patients in need as compared to privileged patients and wealthy corporations. Janiesch et al. [21] summarised the main features of Machine Learning and Deep Learning, to get a better understanding of current intelligent systems. They have explained how analytical models are built through ML and DL. And what are the challenges faced in their implementation. Further, the main issues in Machine-human interaction are also highlighted. The various experts in international PCOS network have proposed a evidence based guideline that was published in 2018. The paper [22] by Kathleen M. Hoeger et al. was brought up in 2021 suggests that it is a priority to translate this guideline to medical practice. They have summarized some of the main points of diagnosis and treatment and also the areas of controversies and challenges are highlighted that can further guide the clinicians and researchers.

Zhou et al. [23] presented a survey article that focussed on the use of Deep Learning in addressing some of the issues in medical imaging tasks; highlighting clinical as well as technical challenges. Also some case studies like abdominal imaging, chest, brain, cardiovascular imaging, digital pathology etc. are mentioned showing the use of various Deep Learning techniques. As a future possibility, these medical images can be combined with other clinical descriptors like blood tests, medications, genomics etc. For this, a good level of privacy and security setup should be built between the hospitals and academic research institutions. Inan et al. [24] have used XGBoost algorithm to classify PCOS and non- PCOS patients. Their main objective was to increase the minority samples and remove outliers in the dataset that could help in solving the class imbalance problem. It is advised for extensive hyperparameter tuning and improved feature selection for better performance.

Peña1 et al. [26] explained the challenges faced when diagnosing PCOS specially in adolescents. Guidelines given by international healthcare professionals promote accurate and timely diagnosis of PCOS, so that consistent care can be provided. The focus is also on improving the diagnostic accuracy and preventing over diagnosis. Boyanapalli et al. [27] performed a brief analysis of various techniques in deep learning that can be used for detecting cysts in ovaries to be benign or malignant. This can predict ovarian cancer at an early stage using ultrasound images. The study can be further extended using MRI or CT scan images and using preprocessing techniques to improve the accuracy.

Soucie et al. [28] tried to analyse the factors affecting PCOS diagnosis delay in the Canadian Health Care System. First, the researchers took the consent of the patients to participate in this study. They, then constructed a diagnosis timeline to help in recalling specific events and dates throughout their diagnosis journey and documenting their experiences before, during and after the diagnosis. The participant's viewpoints were then recorded about the changes in the medical system diagnosis over time. The results concluded on some important points like symptoms in adolescents could not be recognized early, cautiousness about treatment options, uncertain future and creating self-awareness. Certain limitations were found that the data used was of women residing in a particular area and of a particular age group. Another point noted was that it was difficult for participants to recall the diagnosis journey, which could have affected the results. Thomas et al. [29] developed a hybrid system with Naïve Bayes and ANN that could predict the occurrence of PCOS before the condition gets worse. The PCOS classification evaluation is then performed on evaluation metrics like precision, recall, F-measure and specificity. The paper concluded by stating that hybridization could produce better results as compared to individual evaluation.

Nusinovici et al. [30] have compared the performance of different Machine Learning algorithms with Logistic Regression to predict the risk of cardiovascular diseases, kidney diseases, hypertension and diabetes. The study was conducted on 6762 Asian adults considering different percentages of incident cases of these diseases at 6 years. As a result, it was observed that Logistic Regression outperformed other ML models like a single hidden-layer neural network, Random Forest, K-Nearest Neighbor, Support Vector Machine etc. in predicting the risk factor of chronic diseases considering some simple clinical parameters. Zhang et al. [32] have used follicular fluid in their studies to characterize the two groups as

PCOS and non-PCOS. The machine learning models used were based on k-nearest neighbour (KNN), random forest and extreme gradient boosting (XGBoost). They have used Raman spectroscopy in their models, which also proved to be of advantage in detecting changes in a person's metabolic profile. Also the results were found to be of higher accuracy when follicular fluid was used instead of plasma samples.

Elmannai et al. [33] have proposed a model for early detection of PCOS. They have used 541 instances of the dataset collected from Kaggle repository. First the feature selection is performed on this dataset. Then, various ML models (like Naïve Bayes, Random Forest, Support Vector Machine, Logistic Regression etc.) are combined as the base learners to produce the Stacking Ensemble Model. These stacking ML and other ML models are compared based on different Evaluation Metrics like Precision, Recall, Accuracy etc. The ultimate objective of this study is to help healthcare systems in early detection that can reduce long-term complications of the disease. Mathur et al. [34] have suggested a deep learning method S-NET that can simultaneously perform segmentation and 3D TVUS (trans vaginal ultrasound) of follicle volumes in the ovary. This method proved to be helpful in the sense that it ensured the segmentation mask used was consistent and smooth in and did not need any post-processing in 3D space. It reduced any run time errors and also limited the amount of memory needed. In the future, the authors have suggested automatic longitudinal tracking of follicles during IVF cycles, which can help clinical practitioners in deciding the dosage of gonadotropins given to the patients based on the response of follicles. Bharati et al. [35] focused on a data-driven approach to the datasets freely available in kaggle repository. They performed a ranking on the features and found out that the first 10 highly ranked features were the best to predict PCOS. Various classifiers such as Random Forest, Gradient Boosting, Logistic Regression and Random Forest Logistic Regression (RFLR) were then applied on the dataset. The results showed that RFLR demonstrated better accuracy and recall factor. As a future work, these results can be worked out on some different datasets or some hybrid classifier can be produced to improve the performance.

Kodipalli et al. [36] have proposed an early detection and prediction model that could help both in estimating the probability of having PCOS and its other related mental health issues. A fuzzy Technique for Order of Preference by Similarity to the Ideal Solution (TOPSIS) is compared with SVM algorithm by evaluating them on the same dataset. Fuzzy TOPSIS was found to give a 98.2% accurate result while SVM was found to be 94.01% accurate. It was concluded by considering an evaluation of the psychological well being of a woman in the treatment protocol of PCOS. Zhang et al. [37] have tried to identify malignant and benign cysts using ultrasound images fed to pretrained GoogLeNet convolutional neural network, and the rotational invariant uniform local binary pattern (ULBP) features were extracted from each of the images as the low- level texture features. Then these features were normalized and descended together as one feature. Then this fusion is sent to Random Forest Classifier to classify the images as malignant or benign. The low point of the paper is that if the image quality is not good, the algorithm will not work. Also there is a scaling problem that has been left as the future aspect. Harikiran et al. [38] presented a method for detecting follicles in the ultrasound image of ovaries using an Adaptive data clustering algorithm. This algorithm generated accurate segmentation results with simple operation and avoided the interactive input K (number of clusters) value for segmentation. After segmentation, the region properties of the image were used to identify the follicles in the ovary image. The proposed algorithm was then tested on sample ultrasound images of ovaries for identification of follicles and with the help of region properties, the ovaries were classified into categories like, normal, cystic and polycystic ovary with their geometric properties. Ramamoorthy et al. [39] tried to identify the best filter for pre-processing ultrasound scan images and to monitor the growth of cysts in ovary before and after clinical treatment for PCOS patients. The system consisted of two modules namely the Image Pre- processing module and the Image Registration modules.. Image registration is a technique used to spatially align two or more images of the same or different modalities. Affine transformation was used in the final image registration for optimizing the process that comprises translation, scaling, similarity transformation, rotation, shear mapping and compositions of them used in any combination or sequence. It helped in adjusting the spatial information between the source and target image.

Peiffer-Smadja et al. [40] provided a summary through their literature review to address clinicians about the use of ML for Clinical Decision Support Systems (CDSS) in detecting infectious diseases. The introduction part shows a comparison between Expert Systems and Machine Learning, their process of collecting data, translating that into rules and proceeding further with disease prediction. Different summary tables show various ML CDSS used, their requirements of operation and global challenges. Although present ML tools cover a wide variety of clinical outcomes like selection of antibiotics, prediction and diagnosis of diseases etc., but their major limitation is the use of open access dataset and limited number of structured patient variables on which they rely. For the future, the reviewers suggest that the systems to be developed should consider diverse health settings, and the outcomes reported should consider routine clinical care. Denny et al. [41] collected real-time data of patients captured through radiology investigation during their visit to the healthcare center. PCA transformation was used to identify eight main features that proved to be promising in discriminating PCOS from non-PCOS patients. Furthermore, various Machine Learning classification algorithms like Naive Bayes, K-nearest neighbor (KNN), Random Forest (RF) classifier and Support Vector Machine (SVM) were used and their comparison table showed that RF classifier performed better identification. For further indication, the authors have suggested a study on the effect of Vitamin D on PCOS patients, the impact of PCOS on conditions like preterm labor/abortions, and finding out patients with lean PCOS.

Omar et al. [42] developed a prediction model to predict Autism traits in different age groups like 4–11 years, 12–16 years and 18 years and above. The dataset used was the AQ-10 dataset. Initially, the Decision tree -CART algorithm was implemented which was further improved into Random Forest -CART(Classification and Regression Tree) and finally much better results were obtained by merging Random Forest and ID3 classifier. Another major contribution of this work was the development of a mobile application for end users, that could guide them at an early stage with the symptoms, thus reducing the cost associated with delayed diagnosis and preventing the situation from getting worse. One limitation that could be assessed was the marginal performance reported by the model in terms of accuracy when it came to analyzing real datasets. Toosy et al. [43] have presented a review that summarises cardinal traits and generated some treatment algorithms for a less common phenotype of PCOS which is known as lean PCOS. They have provided some management algorithms for this phenotype which include certain lifestyle modifications like dietary changes and regular exercise. Also some medical treatments are suggested that can be followed in consultation of a doctor for conditions like Hirsutism, menstrual dysfunction, acne, infertility etc. The reviewed summary is giving insights of various Machine Learning algorithms [44, 45] used to develop decision support systems for healthcare applications, thus reducing the research gap between healthcare and Machine Learning. A brief analysis of various ML algorithms used along with the techniques for diagnosing various diseases like heart diseases, diabetes, and breast cancer has been shown in a tabular manner. It mentions the datasets used by the techniques and their accuracy in predicting each of these diseases. The results of this review from earlier studies show that Naïve Bayes is 86% accurate in diagnosing heart diseases, SVM gave 96.4% accuracy in breast cancer detection while CART (Classification and Regression Tree) shows 79% accuracy in diabetes diseases. The researchers in [46] and [81] deal with ovarian cysts. While the first one states how cysts are surgically managed if they become malign in post-menopausal women. The most common way stated is laparoscopy but, in the cases, where cysts present some infectious and harmful symptoms and even extra ovarian dissemination occurs, and then laparotomy treatment is considered. The technique is to be applied in treating such systems by taking into account the patient's will and the surgeon's experience. The second one calculates the fluid attenuation values in cysts, based on which it identifies benign and malignant parameters in the cysts.

The authors in [47, 48, 52, 64] have worked on ultrasound images. Each of these have used different methods like gray scaling, histogram equalization etc. for cleaning the data. After that some of the important features were extracted from the images and based on this extraction classification of ovaries into PCO or non-PCO was done. Finally the accuracy of the model was calculated. For future work, it was suggested to generalize the work on all types of cysts including Dermoid cysts.

Ravishankar et al. [49] have demonstrated the use of deep transfer learning in situations where training data is limited, specially in problems related to medical images. Ultrasound images of kidneys were collected from GE Health care LOGIQ E9 scanner to perform image classification. For this Convolution Neural Network(CNN) is first trained using Image Net, and then transferred to perform the classification task. The model is then hybridized using the techniques of transferring and tuning to achieve 20% better performance in classification of kidney detection problems.

Shan et al. [50] have performed a study on Li people to analyze the risk factors that could lead to PCOS. This analysis would help in early diagnosis and treatment for it. Questionaire method was adopted for data collection of Li patients. Then they performed analysis like univariate and multivariate. The multivariate analysis treated PCOS as dependent variable and other risk factors like menstrual cycle disorder, infertility or mother's irregular cycle, family history of clinical conditions like diabetes, lack of physical exercise etc. Further, these risk factors were confirmed using multivariate logistic regression. Setiawati et al. [51] have proposed an approach for clustering images for follicle segmentation using Particle Swarm Optimization (PSO) method. To generate a more compact cluster technique like Mean Structural Similarity Index (MSSIM) and Normalized Mean Square Error (NMSE) were used. Also the authors have evaluated that contrast enhancement, if does not precede PSO image clustering, can produce a close Region of Interest (ROI). They have put forward the use of Logistic Regression Classifier to classify the extracted follicle features automatically.

Witchel et al. [53] give an overview of the pathophysiological factors affecting PCOS and various treatments related to it by giving emphasis on adolescent girls. The authors explain that a Rotterdam criterion is acceptable for adult women, but in case of adolescent girls there are different diagnostic criteria that can be used instead of pelvic ultrasounds. These include looking for menstrual irregularity, clinical signs like increased androgen levels etc. Various methods to manage PCOS are also mentioned like education and counseling, lifestyle modifications, treatment of acne and hirsutism etc. This comprehensive analysis on adolescent girls will help in identifying such girls who can then be given timely implementation of planned transition methods. This would also help in improving their quality of life.

Zhu et al. [54] have explained whether a direct association exists between PCOS that is predicted genetically and other diseases like type 2 diabetes, coronary heart disease (CHD) and different subtypes of strokes. They have taken samples of Asian and European males, females and both together. The methods used in the analysis were MR analysis using IVW, weighted – median and MR-Egger. In all cases it was observed that there is no direct influence of genetically predicted PCOS and the diseases mentioned. The authors have also clearly mentioned that other high-risk features of PCOS like obesity, elevated testosterone etc. may better explain the association between PCOS and cardiometabolic diseases. So for further study, suggestions are made to give more emphasis on women specifically with high risk features of PCOS rather than all women with PCOS.

de Matos et al. [55] have presented a review of all the ML techniques that could be used to analyse Histopathological Images (HIs) for cancer detection problems. According to this study, analysing these images can sometimes lead to disagreements even among the experienced pathologists. Thus, this review shows the various Computer Aided Design (CAD) systems that can accelerate the task of analysis. The various ML methods studied here include segmentation, feature extraction, shallow methods and deep methods. A summary of the list of publications arranged chronologically along with the tissue/organ studied using each of these methods is presented tabularly. The review has been concluded with an increase in demand for Deep Learning methods as they can deal with raw His with little or no preprocessing; while it is found that there is a decline in the use of segmentation and feature extraction.

Simons et al. [56] have tried to resolve the uncertainty of whether PCOS is the cause behind the increased risk of Coronary Artery Disease (CAD) or obesity, a major factor behind PCOS results in CAD. They have collected female-specific data from different sources. Inverse- variance weighted (IVW) method was used to perform primary analyses while the Simple Median, contamination mixture analyses and penalized weighted median were carried out to generate results that were quite robust. All the outcomes led to the same conclusion that there is no direct relation between the genetically predicted risk of PCOS and the risk of CAD. Instead, PCOS and CAD are both influenced by obesity predicted genetically and its effect on metabolism.

Prasanth et al. [57] have used various Machine Learning classification models to predict early risk of diabetes. As diabetes is the cause of various diseases, the earlier it gets detected, the better it can prevent the spread of other diseases. The dataset is first analysed and feature selection is performed. After that various classification algorithms like KNN, SVM, Logistic Regression, Gradient Boosting etc. are evaluated for their performance. It is found that Gradient Boosting gave better accuracy. Jusman et al. presented [83] the comparison of the performance of the Multi Layered Perception and Radial Basis Function Classification that are applied on of Lung Cancer Data. Ricciardi et al. discussed [84] about the prediction of the coronary artery disease by applying the Linear discriminant analysis and principal component analysis. Wang et al. [87] have analysed 143 cases of COVID 19 patients to study the severity of the disease and relation between clinical and biochemical factors. In addition to age, other factors like albumin, C-reactive protein, D-Dimer etc. also affect the occurance and severity of the disease. The researcher [90, 91, 108] have proposed some Machine Learning applications for predicting neurodegenerative disorders and heart diseases. In the first one, the importance of feature extraction is shown, classification and then clustering is performed after feature extraction is done. While the second one creates a hybrid model using several different classifiers, first performing feature selection on heart dataset and then after creating the model evaluating their performance metrics using Area Under Curve (AUC). The third work tries to compare different classification models again working on heart dataset and trying to find out the best model.

Azziz et al. [58] have tried to prove whether Rotterdam Criteria 2003 should be considered for the diagnosis of PCOS rather than NIH 1990. He has mentioned the two new phenotypes introduced by Rotterdam Criteria 2003. These include first, conditions like ovulation in women with PCOS and hyperandrogenism and second, oligo-anovulation in women with PCOS but without the presence of hyperandrogenism. He has also stated the risks and benefits of considering Rotterdam 2003 criteria as PCOS. It was finally concluded that the two new phenotypes identified are unclear and their fertility rates are very few. So, further studies are needed to confirm these findings. Hart et al. [59] have tried to find out how many women who had PCOS were admitted in hospitals as compared to those without PCOS. This is analyzed on a Western Australian population. The results were analyzed on number of women diagnosed with PCOS, their reproductive health, non-reproductive health which includes conditions like diabetes, obesity and other nutritional and metabolic diseases. Cancer health, cardio-metabolic and thrombotic health and mental health were the other areas that were studied.

To overcome the problem of inefficiency in results, low variability in manual detection of number and size of follicles, Adiwijaya et al. [60] have suggested automatic detection methods. In first scheme, the real follicles were segmented from homogenous region. In second method, the segmented region attribute was measured as the follicle by counting the number and size in USG images. This then helped in categorizing the image as PCOS and non-PCOS. The region growing methods that included region-based and seed–based identification were used. The Euclidean distance method was further used to quantify the follicle. Sayyah-Melli et al. [61] have tried to determine what psychosocial factors affect women with PCOS. They have formed two groups- one of women with PCOS and the other without PCOS. This data was collected using a questionnaire method. The Primary Care Evaluation of Mental Disorders and Patient Health Care (DSM-IV) helped in analyzing the major psychopathologic disorders and other syndromes. Further, a clinical psychiatrist verified the results. Overall it was concluded that depression and anxiety were the most common disorders in PCOS patients, they were less educated and suffered from low socioeconomic status. It is suggested that proper counselling of these patients should be done to maintain their health and care. Jolliffe et al. [62] have shown the importance of a feature selection technique known as Principal Component Analysis (PCA). Its variants and application areas are also discussed. It tries to find uncorrelated variables that maximize variance, thus making PCA, an adaptive data analysis technique.

The main objective of Gibson-Helm et al. [63] is to find out the diagnosis experiences of PCOS in women, the concerns that they have shown after that and the satisfaction `level with the information provided to them. The data was collected from different regions of the world like North America, Europe etc. through a questionnaire circulated through websites of PCOS support organizations – PCOS Challenge (USA) and Verity (UK), social media or even mailed to women. Univariable and Multivariable Logistic Regression was used to find out the relation between satisfaction level of diagnosis experience and other contributors like number of health professionals visited, time to diagnose, age, region of residence etc. It was concluded that major concern that needed lifestyle changes among women was excess weight management. Also, women are not satisfied by the information provided through educational resources and their reachability. Some initiatives could be taken by the researchers in future to prepare some guidelines that can not only educate but improve the diagnosis experience as well. Rihana et al. [64] have used SVM classifier to detect cysts and classify them in the ovary ultrasound images. The images are first pre-processed and then feature extraction is performed from the defined region of interest. Finally, the classification and validation is performed using ROC. For better evaluation, the database sizes can be increased. Deep learning models [65, 66, 79] and CAD systems can be designed for detecting different types of cancers like ovarian, liver and breast cancer. Some of them use ultrasound images, others have collected dataset from TCGA or Mammographic Image Analysis Society (MAIS). The models are then compared on different performance metrics to evaluate their efficiency. Engmann et al. [67] in their work try to show the importance of racial and ethnic differences in metabolic phenotypes among women suffering with PCOS. They have collected data from a randomized controlled clinical trial PPCOS II based in US. It was observed that the parameters were not sufficient to evaluate this and different thresholds were needed for cardiometabolic markers to detect this. Similarly the work in [80] reflect how people with different ethnic groups are affected differently with PCOS. Mohammad et al. [68] presented a review to demonstrate what factors determine the presence of PCOS, especially in adolescent patients, where there is an overlap between PCOS features and occurance of puberty. One of the main factor that is analysed here is AMH, synthesised from ovaries and also seen in ultrasonographic images. Dewi et al. [69] have designed a system that can help in classifying ovaries as PCO/ non-PCO by using methods used for extracting features and Competitive Neural Network. They used ultrasound images which were first pre-processed, and then the features were extracted using the Gabor Wavelet method from the segmented image. Finally, CNN is used to classify the ovary to obtain the classification result as PCO/non-PCO. With this system an accuracy of 80.84% is achieved using 32 feature vectors.

Balogun et al. [70] have tried to predict the occurrence of infertility in women using some of the supervised machine learning approaches like– Naïve Bayes, multi-layer perceptron and Decision tree. They have collected a dataset of 39 patients from a hospital in Nigeria. The gynecologist helped them in identifying the 14 variables necessary for classifying the women as infertile. For training the model, the method used was tenfold cross validation used and the results were analyzed using WEKA tool. The confusion matrix was used to demonstrate the overall accuracy which proved that Multi-layer perceptron is an efficient algorithm to help in predicting infertility. It is suggested to increase the number of records for more accurate results. A relationship plot can be generated to show the effect of selected variables on the likelihood of infertility. The researchers also [71, 73] provide some of the risk factors and guidelines or recommendations to detect ovarian and adnexal abnormalities. They also provide some practice points to be prioritized to improve the health outcomes of women suffering with PCOS. Some of the changes in the guidelines reflect on improving education, focussing on lifestyle changes, providing more accuracy in the process of diagnosis etc.

Wolf et al. [72] have tried to come up with a review that determines the existence of PCOS in women across different geographical locations, racial or ethnic groups. They have noticed that due to the use of different diagnostic criteria like Rotterdam 2003, NIH 1990, AE-PCOS 2006, there are alterations in the results. Also, if all 3 criteria are combined, prevalence rate of PCOS reduces to as low as 1.6%. They have also focussed on a particular reproductive age group of women like 18–45 years. So they have concluded that there is no significant difference in the existence of PCOS across different geographical, ethnic or racial groups because the diagnostic criteria used by them are different. A better diagnostic criteria is suggested to be used, through which less individuals are left undiagnosed.

Ji et al. [74] have tried to compare the effect of modulation in Autonomic Nervous System(ANS) in women with and without PCOS. For this, they have collected the data of Heart Rate Value (HRV), Body Mass Index (BMI), and physical reports of women with PCOS and those with regular menstrual cycles. It was observed that women with PCOS showed higher values of low – frequency (LF) power, normalised low-frequency (norm LF) power and LF/High Frequency (HF) ratio as compared to the control group. So, ultimately they have concluded that sympathetic modulation may increase in women with PCOS. The study could be extended further to test with BMI and BP matched control groups. Sarker et al. presented [75] a review that indicates that the clear shift of the techniques using Artificial intelligence apply in medical area with the methods of the deep learning.

Sreejith et al. [76] have tried to develop a system for Clinical Decision Support that can help physicians in examining PCOS. They have used Red Deer Algorithm (RDA) for choosing the features and Random Forest (RF) classifier to segregate the dataset. After training the dataset, the testing is performed on selected feature database and performance is evaluated by comparing it with other classifiers like Naive Bayes, SVM, k-NN, LR etc. Also the proposed method is then compared with other Wrapper methods using GA, PSO, ACO etc. In both the comparisons it is observed that this proposed method demonstrated a higher accuracy of 89.81. In future researches, the scalability of the model can be evaluated using other datasets or the same dataset can be used with other optimization techniques. Barber et al. [77] presented a review to sightsee the impacts of weight gain and obesity on the pathogenesis of PCOS. Maadi et al. [78] presented a review where importance of human and AI interaction is shown in ML applications. To prove this fact, some questions were designed and were answered by analysing how humans are involved in developing ML applications, how algorithms are designed in stages- like producing data, preprocessing it, building a model and evaluating and refining it. It is concluded by offering opportunities for future research to explore more in this human – AI interaction.

Sendur et al. [80] also discussed about the Influence of the ethnicity on the various characteristics of the polycystic ovary syndrome. Dapas et al. [82] have used a genome wide association study (GWAS) of PCOS published previously to find out the subtype specific genetic associations of the subtypes of PCOS that were reproducible. The technique they used was unsupervised hierarchical clustering which identified two subtypes of PCOS. One of the subtype defined was reproductive while the other one was a metabolic group. Each of these subtypes were characterised by some traits or features. The study concluded that the subtypes identified appear to possess different genetic architectures. The limitation of this study is that only European ancestry PCOS causes are included.

Munjal et al. [85] have demonstrated the use of a Machine Learning approach which is Genetic Algorithm. The algorithm works by selecting the major attributes i.e. the signs and symptoms of PCOS. These attributes are then given as input to three different types of classifiers namely Decision Tree, Random Forest and Extra Trees. These classifiers are ranked based on different accuracy parameters like F1 Score, Precision, Confusion Matrix etc. to decide which one best classifies the diseased and other patients. The study can further be extended by selecting the genetic attributes from a large number of datasets thus helping a larger community in the diagnosis process.

The work by Prapty et al. [86] tries to help women who are affected by PCOS to monitor themselves by noticing the key features identified using a decision tree. They have applied different machine learning approaches and then constructed a decision tree based on the best-performed classifier. the classifier identified here as a good competitor is Random Forest as it shows less variance, is more flexible and also solves over-fitting problem. They have also tried to verify the attributes chosen with Principal Component Analysis (PCA) and were satisfied by the principal components chosen.

Janssens et al. [88] demonstrate that how the ROC curve is a different method to present risk disseminations of diseased and non-diseased individuals and how the form of the ROC curve notifies about the overlap of the risk dispersals. The objective of the work done by Nandipati et al. [89] is to identify which features and classification model best help in the prediction of PCOS using two tools namely Python – Scikit Learn package and Rapid Minor. They have used five classifiers, two ensemble classifiers and some feature extraction methods to make the comparison. The execution of both the tools is analysed using all the classifiers and taking all the forty attributes together. After that the algorithms are compared using ten and twenty four common selected features. The paper concluded that regardless of the different tools used Random Forest outperformed other algorithms. Also it is analysed that the performance of the tools depends on the kind of the dataset used and pre-processing steps followed. Escobar-Morreale. [92] has provided a clear definition of PCOS, its symptoms and its subtypes like Classic PCOS and Ovulatory PCOS. He has used a term heterogeneous for PCOS which means that the various phenotypes which affect it, the clinical methods and its metabolic consequences are multiple in natures. Further, some environmental factors like unhealthy diet and other physical habits can exaggerate the situation. So it is suggested here, that, if the person shifts to certain lifestyle changes, then at least the non-genetic inheritance of PCOS can be prevented. At the end it is concluded by suggesting some clinical methods for diagnosing PCOS. Prajapati et al. [93] discussed about the creation and Evaluation of a Polyherbal Tablet for Polycystic Ovarian Syndrome.

Armanini et al. [94] have proposed a review that discusses about some controversies in diagnosing and treating PCOS. Also pros and cons related with some contraceptive therapies are mentioned. According to the study the controversies in treatment mainly arise due to the different definitions related with various features of PCOS and thus different criteria are practiced by different clinicians for their diagnosis. At the end it has been concluded that there is no particular treatment to PCOS. The therapy varies from patient to patient by also considering their effects on metabolic and cardiovascular features. The review also states that the recent guidelines given internationally have first recommended some modifications in lifestyle and dietary patterns as the first treatment for all women suffering with PCOS. Armstrong et al. [95] in their study have tried to evaluate whether Serum Anti Mullerian Hormone (AMH) can predict four different phenotypes of PCOS – i.e. phenotype A, B C and D; all with different combinations of traits like amenorrhea/oligomenorrhea (AOM) hyperandrogenism (HA) and polycystic ovaries (PCO). 227 women were included for this study that was diagnosed with PCOS and the rest 103 women did not have PCOS. Their AMH levels were checked using Beckmann Access 2 and also their age, AMH and BMI were analysed using a different method. They concluded that AMH is a strong prognosticator of different phenotypes of PCOS. This aspect can be taken by future researchers in their analysis. Nssibi et al. [96] presented an in-depth inspection of nature-inspired metaheuristic approaches for the feature selection problem, with an emphasis on representation and search algorithms. Spremovic et al. [97] have presented a review on the presence of non-alcoholic fatty liver disease (NAFLD) as a major reason for severe liver diseases in women affected with PCOS in the western part of the world. According to the references given in this article, the liver diseases are linked with some major factors like obesity, insulin resistance, inflammation etc. Some of the scientific studies that show the spread and control of the diseases and their risk factors are also mentioned. Ultimately, they have concluded that there is insufficient knowledge about the reasons behind the development of NAFLD in PCOS women; Future research is needed in this area to identify the true relationship between the two.

Kaur et al. [99] have tried to design a framework that helps in classifying the food images and suggested the list of food items to the patients suffering from PCOS. They have used pre-trained CNN architecture (i.e. variants B0-B7 of Efficient Net model) which is further fine-tuned by adding four additional layers. The model gives the calories, fats, proteins, fibres and carbohydrates present in the given food image intake by a person. The balanced calorie requirement needed by a PCOS woman is then calculated using K-Means clustering and Random Forest. This helps in predicting clusters of food items needed as per the required nutrition content.

Nazim et al. [100] have tried to demonstrate a novel method for predicting PCOS. The main focus of this early prediction is to prevent the women from severe health complications. First, they have selected the best features from the clinical and physical parameters based on the optimised chi-squared (CS-PCOS) mechanism. Using this, they have further proposed Gaussian Naive Bayes (GNB) and compared it with other nine machine learning algorithms. They have concluded that the features like prolactin (PRL), thyroid stimulating hormone (TSH), blood pressure and pregnancy are the major predictors of PCOS. In future, the dataset can be extended and some more data balancing techniques can be used, also deep learning techniques can be applied for PCOS prediction. Panicker et al. [101] have used a self built CNN model for analysing the presence of PCOS from ultrasound images. First, they have segmented the follicles in ovary ultrasound images using filters of CNN, and then classification as PCO and non-PCO is performed using the fully connected layer present in CNN. They have concluded their findings that the model showed an accuracy of over 83%.

Hosain et al. [102] have recommended PCONet – a type of CNN to detect PCO from ultrasound images of the ovary. Also they used another model i.e. InceptionVR, which is a pre-trained CNN consisting of 45 layers. It is fine tuned by transfer learning method to classify the same ultrasound images as done by PCONet. They have then compared the performance of both these models on various parameters to and the results show that PCONet showed superior accuracy. Kumari et al. [103] have endorsed SMOTE- stacked hybrid model for early detection of PCOS. SMOTE stands for Synthetic Minority Oversampling Technique which is used in balancing the dataset. The ensemble technique Stacking is used to produce a hybrid model with six different classifiers namely Logistic Regression (LR), Naive Bayes, Decision Tree (DT), Support Vector Machine (SVM), Random Forest (RF), AdaBoost (AdaB). At the base level, the individual classifiers are first trained, the predictions generated at the base level is utilised for training at the meta level and six hybrid stacked models are produced. All these models are evaluated for performance and it is assessed that Stack-AdaB showed promising results. Harshvardhan et al. [104] prepared a study where they tested different generative models used in AI like Gaussian Mixture Model, Hidden Markov Model etc. Each of these are studied and implemented to provide readers some guidance while deciding which model to choose when dealing with a problem. Smiti et al. [105] focussed on importance of analysing patient's data in the medical field. They have shown how ML can prove to be an effective tool in data analytics by initiating with concepts of medical field and Machine Learning. Further, they presented some challenges to be studied, so that an effective medical diagnosis solution is obtained. The different works by different researchers [106, 107, 109] present different Machine Learning and AI tools to analyse some of the diseases and terms related to healthcare, for example, membroanaytic anticancer peptides, type 2 diabetes and Alzheimer's disease.

Table 1 gives a summary of some of the important findings from the research analysed so far, giving emphasis on the focus area mentioned in the papers.

3.1 Result analysis of existing work

In this study, we have tried to analyse different Machine Learning algorithms used by different researchers in identifying PCOS problems. The researchers have used different

Table 1 Related survey on PCOS problem

References	Release Year	Coverage Period	Comprehensive Analysis	Future Directions	Algorithm/Tool used
Kumari et al. [103]	2023	2017–2021	Moderate	Not Available	Early detection of PCOS using SMOTE – Stacked hybrid model
Sreejith et al. [76]	2022	2015–2022	Moderate	Available	To propose a clinical decision support system that can classify PCOS using Random Forest(RF) classifier based on the features selected using Red Deer Algorithm(RDA)
Mehr et al. [6]	2022	2011–2021	Moderate	Available	To predict and analyse PCOS by applying different machine learning classifiers with whole set of features and reduced set generated using feature selection methods
Tiwari et al. [7]	2022	2015–2022	Extensive	Available	Evaluating different machine Learning Models based on their threshold value and selecting the best model for PCOS diagnosis
Ahmed et al. [8]	2022	2013–2021	Moderate	Available	Predicting Uterine Fibroids using VGG016 Classifier and comparing it with other technologies given in its Literature survey
Rathod et al. [9]	2022	2015–2022	Moderate	Available	To compare different machine learning classifiers that can help women in predicting whether to seek medical help for PCOS or not
Zigarelli et al. [10]	2022	2004–2022	Extensive	Available	To create self- diagnostic prediction models for evaluating PCOS status
Suha et al. [98]	2022	2015–2021	Extensive	Available	To propose a hybrid model that could combine traditional classifiers in machine learning with deep neural network for feature extraction. XGBoost was used as a metalearner using ensembling for image classification
Nasim et al. [100]	2022	2012–2022	Moderate	Available	To extract important features using Chi-Squared (CS-PCOS) mechanism on the clinical parameters related to PCOSand then classify the data into PCO or non-PCO using a comparative study of proposed Gaussian Naive Bayes(GNB) with other machine learning classifiers
Panicker et al. [101]	2022	2009–2021	Moderate	Not Available	To classify the ovaries as PCO or non-PCO using USG images by the use of full connected layers present in CNN
Hosain et al. [102]	2022	2012–2021	Moderate	Not Available	To detect polycystic ovaries using PCONet- a type of CNN. Also to compare its accuracy with another type of pretrained CNN i.e. InceptionVR on the ultrasound images
Gopalakrishnan et al. [15]	2021	2011–2021	Moderate	Available	To calculate the performance of different classifiers on extracted features to detect follicles and classify PCOS

Table 1 (continued)

References	Release Year	Coverage Period	Comprehensive Analysis	Future Directions	Algorithm/Tool used
Inan et al. [24]	2021	2011–2020	Moderate	Not available	To increase minority samples and remove outliers in the dataset in order to analyse class imbalance problem. The selected features are fed into XGBoost algorithm to classify PCOS and non-PCOS patients
Boyanapalli et al. [27]	2020	2014–2020	Limited	Not Available	A contrasting study of different methods mentioned in the literature for detecting Ovarian Cancer
Thomas et al. [29]	2020	2009–2018	Limited	Not Available	Calculating the Performance Metrics of algorithms like Naïve Bayes, ANN and their hybrid classifier and comparing their results based on some of the performance metrics
Bharati et al. [35]	2020	2011–2020	Limited	Available	Calculating the accuracy of different classifiers used in PCOS classification (using only 10 features) and comparing them with the models used in the literature review
Prapty et al. [86]	2020	2002–2011	Moderate	Not Available	To help women monitor their PCOS level by the use of decision tree built on the basis of key features identified by different Machine Learning Algorithms
Zhang et al. [37]	2019	2007–2018	Moderate	Not Available	Using Random Forest Classifier to categorise the images of cysts in ovary as Benign and Malignant
Denny et al. [41]	2019	2006–2018	Moderate	Available	Calculating Accuracy, Sensitivity, Specificity and Precision of various models used in PCOS classification
Dewi et al. [69]	2018	2007–2015	Limited	Not Available	To group the ultrasound images of ovaries as PCO or non PCO using Competitive Neural Network
Wisesty et al. [47]	2017	2006–2015	Limited	Not Available	To preprocess the ultrasound image, performing feature extraction and categorisation of PCOS using modified form of Back propagation Algorithm
Cahyono et al. [48]	2017	2004–2016	Limited	Available	Classifying Ultrasound images into PCO and non-PCO class using CNN classifier

parameters like Precision, Accuracy, F1 score etc. to understand the effectiveness of their algorithms.

The work done in [6, 7, 35, 41, 76, 86, 89] have worked on structured dataset and showed that Random Forest outperformed other algorithms in detecting presence of PCOS. Some of the researchers in [6, 29, 103] have performed ensemble learning and hybridization to combine the positive effects of well trained classifiers. In other works like [12, 13, 16, 25, 31, 98], the study has been conducted on different medical images like pathological, endoscopic or ultrasound images. They have applied CNN or DCNN to segment and analyse different parts of the images and classify them accurately.

The literature review carried out here clearly summarises the benefits of using some of the algorithms and how ensembling and hybridization can produce better results as compared to using a model individually. The references also mention about the quantity and quality of data that should be used to make the analysis of results more profound.

4 Research strategy

In order to organise the literature studied in the area of PCOS diagnosis, a systematic method has been followed. The Fig. 3 below provides a diagrammatic approach of the Review Process followed in this paper. It helped in documenting the major points which further led to smooth execution of this Systematic Literature Review (SLR). The major steps included are explained as follows:

- Recognised the necessity for performing SLR in comparison to any other techniques (subtopic 4.1).
- Framed the research questions that account for this motive of PCOS diagnosis (subtopic 4.2).
- Constructed a collection of keywords that can help to inquire about databases which are relevant to our research and find out the publications from the top medical journals/ conferences/ scientific reports/workshops etc. (subtopic 4.3).
- Registered the criteria to include or exclude the papers recovered, in order to refine the set of papers so that they fit in more appropriately in our research scope(subtopic 4.4)
- Tried to relate the relevancy of papers with their citations to support them.
- Framed an initial pathway and filtered it as we proceed further to obtain the set of relevant papers.
- Pointed out some important highlights like comparison between common Machine Learning Algorithms used in healthcare, extracted from the literature that exists in

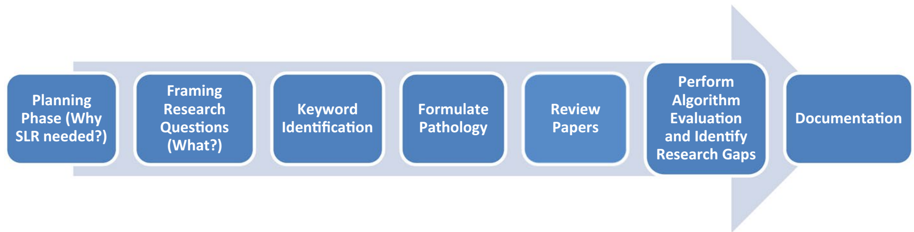

Fig. 3 Systematic literature review process

order to come up with some major findings that could help the entire research community trying to work in the similar direction (subtopic 4.5)
- Recognised some major gaps left in the existing research, trends and the future works based on the study (subtopic 4.6).

4.1 Need of systematic literature review

With the increase in research being carried out in the area of medicine, it is hard to keep a note of the literature already existing. The prime way we can deal with this situation is to conduct a review. The main reason behind carrying out a review is to analyse the different researches to put all the studies into an all-inclusive summary with the purpose of resolving the issues left behind, keeping the main topic into consideration. This review could be a simple survey, or a more complex SLR. SLR follows a pre – defined protocol, which is carried out by selecting and evaluating articles further.

4.2 Research questions

The following Table 2 puts forward the research questions that are considered in performing SLR. These objectives are then carefully analysed and answered so that they can help in framing future objectives for the researchers.

4.3 Keyword identification

4.3.1 Search strategy

The main focus of this paper is to analyse deeply the various Machine Learning Algorithms that have helped researchers in identifying various types of diseases, particularly focussing on Polycystic Ovary Syndrome (PCOS). In order to achieve this, a large set of articles published in this field were retrieved from various sources. For the initial level, an extensive search of articles was performed on the following electronic research databases: Elsevier, IEEE Xplore, Springer and Google Scholar. Apart from the aforementioned repositories, some other sources such as conference proceedings, workshops and symposia were also taken into consideration. The keywords were carefully identified and divided into four sets based on similarity. The first set of keywords contained the major terms which are used to perform classification and clustering. The second set of keywords revolved around the various adjectives related to the problem in hand. The third set itself contained the keywords or algorithms used in structured data analysis. Finally, the fourth set contained the keywords relating to the image identification and performing cluster analysis. Although the major focus in this article is on PCOS identification and diagnosis, a lot of articles also talk the psychological effects like stress, anxiety, mental disorders and physical health issues like obesity, diabetes, cardiovascular and metabolic disorders. Hence, these keywords are included to prevent loss on some major insights that can be drawn from such articles. A careful combinations of various identified keywords were used to apply search queries on the above mentioned research databases.

The keywords identified are mentioned in the below keyword-sets.

Keyword_set_1 = {Machine Learning, Data Science, Deep Learning}

Table 2 Gives an analysis of accuracy of different models used in various research papers

Author/Reference	Model with best performance	Accuracy obtained
Mehr et al. [6], Elmannai et al. [33]	Ensemble RF and MLP; Stacking Ensemble Model	98.89% [6]; 100% [33]
Tiwari et al. [7]; Zhang et al. [37]; Denny et al. [41]; Sreejith et al. [76]; Nandipati et al. [89]	Random Forest	93.25% [7]; 89.02% [41]; 89.81% [76]; 93.12% [89]
Ahmed et al. [8]	VGG 16	98.5%
Zigarelli et al. [10]	Cat Boost	81% to 82.5% without using any invasive measures in the patient models and 87.5% to 90.1% prediction accuracy when both noninvasive and invasive predictor variables were used
Sumathi et al. [31]; Cahyono et al. [48]; Dewi et al. [69]; Suha et al. [98]; Panicker et al. [101]	CNN, DCNN, CNN with transfer learning	85% [31]; 76.36% [48]; 80.84%[69]; 99.89% [98] accuracy of over 83% [101]
Gopalakrishnan et al. [15]	SVM classifier using RBF kernel	93.82%
Thomas et al. [29]	hybrid system with Naïve Bayes and ANN	95%
Bharati et al. [35]	Random Forest Logistic Regression(RFLR)	91.01%
Kodipalli et al. [36]	Fuzzy TOPSIS	98.2%
Rihana et al. [64]	SVM Classifier	90%
Balogun et al. [70]	Multi-layer perceptron	74.4%

Keyword_set_2 = {Poly Cystic Ovaries, Uterine Fibroids, Follicles, ovarian cancer}
Keyword_set_3 = {Classification, Regression, SVM, Random Forest, Decision Tree}
Keyword_set_4 = {Clustering, Image detection, CNN, AdaBoost, VGG16}

It is found that each of these keywords are relevant to our review. The first set of keywords provide an overall area that could be used in Healthcare system. The second keyword set relates to important terms that are used interchangeably to identify problems in a female's ovaries or uterus. The third and fourth keyword set lists the major algorithms that have been already used in classifying and detecting Polycystic Ovaries, fibroids or cancer in the uterus. These keywords could further help the researchers formulate their own algorithms or produce a hybridization of any of these techniques.

4.4 Criteria for including and excluding articles

A series of steps are put forward on the set of research articles derived as a result of the various keywords selected on the above mentioned databases and they are selected based on the criteria to include or exclude them. This criterion is listed in Table 3 which helps us in refining the search process, to be in line with this SLR, and prevent us from diverting from the main focus.

4.4.1 Articles selected for analysis

The publications retrieved in the result set after searching them based on the keywords were very large. This result set was then filtered through various methods or strategies. An explanation of the same is mentioned in Table 4. Initially, the articles which were not related or unconnected were discarded based on the insignificance of titles. After that the criteria for including and excluding the articles were applied to further bring down the set of research papers. As a result of this, 1327 articles were advanced for next study. Out of the remaining, 411 articles were rejected after reading their abstract. The left over 148 papers were read completely, but 73 of those read, were found out to be out of context or not relevant to the topic. Finally, 109 articles were selected for final analysis to be made in detail.

The Fig. 4 shows a pie chart depicting the distribution of Papers based on their Publisher.

Table 3 Research questions identified

Research ID	Questions
RID 1	What are the various Machine Learning Algorithms used in Health Care?
RID 2	What are the strategies already used in healthcare sector?
RID 3	How PCOS is an area of great concern in medicine?
RID 4	What are the various challenges faced by the researchers in this area?
RID 5	What performance metrics should be considered to evaluate the effectiveness of Machine Learning model?
RID 6	What are the gaps left in the literature and future aspects that can be taken into consideration?

Table 4 Criteria for including/excluding articles

Criteria	Included	Excluded
Source of Publication	• Journals • Conferences • Reviews (They are considered for review as they contain articles most relevant to our research topic.)	• Government Publications • Policy Documents • Un-published research (It is difficult to access government publications and policy documents as lot of MOUs are to be considered, and some of them are also not easily available.)
Type of Publication	• Review articles • Researches • Books • Chapters	• Reports • News Articles
Time span considered	2013 – 2022 (The articles in this time span focused on latest information and database. It also covered some updated Machine Learning Developments.)	Before 2013 (The database used had become outdated, also the techniques used in these studies have evolved to a greater extent.)
Language used	English	Non-English
Type of Review	Reviewed by peers	Not Reviewed by peers
Any Other Criteria	• Prime Quality • Comparative Study	• Unconnected Contents • Inferior Quality • Comprehensive text not accessible

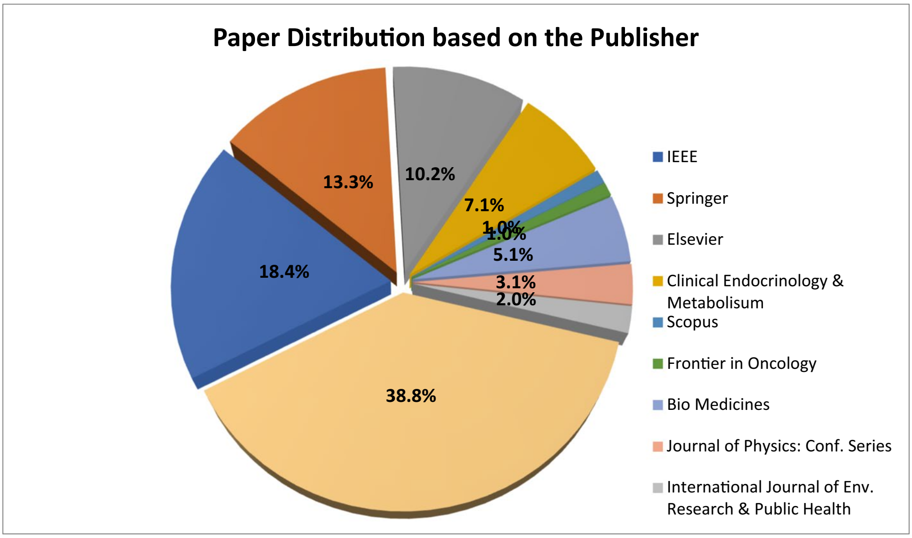

Fig. 4 Paper distribution based on the publisher

The Fig. 5 shows how the Papers are being selected based on their Type. Around 66.3% of them have been taken from well known journals, 25.5% are being taken from conference papers and the rest are from various other sources.

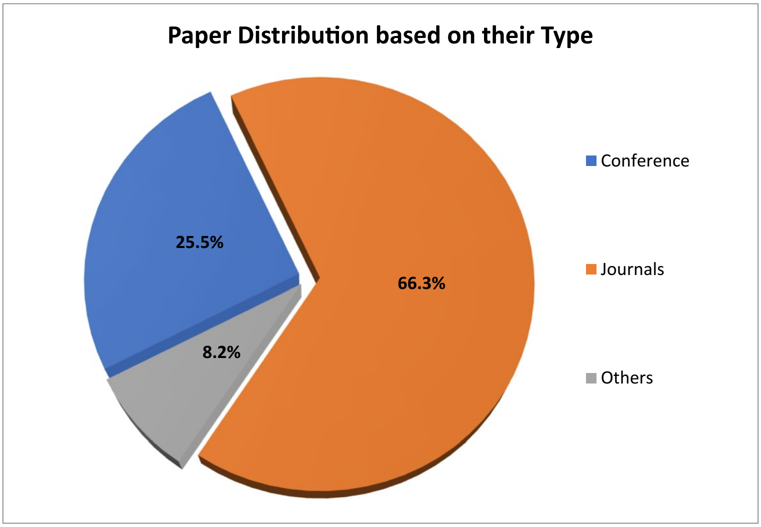

Fig. 5 Paper distribution based on their type

Table 5 No of papers retrieved during article selection

	IEEE	Elseiver	Springer	Clinical Endocrinology	Biomedicines	Others	Total Remaining
Results Based on keyword search	2330	805	504	52	39	85	**3815**
Excluded based on titles	1800	400	250	12	8	18	**1327**
Excluded applying Criteria	335	225	135	15	13	11	**593**
Excluded after Abstract study	175	135	75	7	7	12	**182**
Excluded after detailed study	**8**	**35**	**13**	**11**	**6**	**0**	**109**

4.5 Comparison among some of the common machine learning techniques repeatedly used in Polycystic Ovary Identification

Table 5 below lists the number of papers retrieved while selecting the articles, while Table 6 provides a comparative study of the frequently used Machine Learning algorithms in detecting Polycystic Ovaries. It also lists their advantages and disadvantages.

Table 6 Comparison among the commonly used ML algorithms in PCOS

S.No	Name of the algorithm	References/Authors depicting the use of particular algorithm	Type of Data analyzed by the algorithms	Generalised Application areas(Healt-h sector in particular)	Overall Advantages	Particular Shortcomings
1	**SVM**	Rathod et al. [9], Gopalakrishnan et al. [15], Denny et al. [41], Rihana et al. [64], Kumari et al. [103]	Works well with structured, unstructured as well as semi structured data	Classification of images, Predicting a disease like diabetes, Geo and Environmental Sciences etc	**1. It** works well when there is a clear distinction margin between the identified two classes 2. It performs better in high dimensional spaces 3. Due to the use of small datasets only, it is more memory efficient	1. It does not work well for larger datasets 2. Due to the use of a hyperplane in classification, its probability of predicting the result cannot be clearly explained
2	**Decision Tree**	Rathod et al. [9], Balogun et al. [70], Munjal et al. [85] Kumari et al. [103]	Structured data	Classification and Regression problems like Drug prediction, making an optimal choice like Surgery	**1. I**t does not require scaling and normalization of data 2. Decision tree can be constructed even if there are missing values in the dataset	**1.** It is expensive to build as the training time is high and model computations are complex 2. A slight change in the dataset can change the entire structure of the decision tree thus making it an unstable algorithm

Table 6 (continued)

S.No	Name of the algorithm	References/Authors depicting the use of particular algorithm	Type of Data analyzed by the algorithms	Generalised Application areas(Healt-h sector in particular)	Overall Advantages	Particular Shortcomings
3	**Random Forest**	Tiwari et al. [7], Zhang et al. [37], Denny et al. [41], Kaur et al. [99]	Structured and unstructured data	Predicting diseases like Breast cancer, Diabetes; Predicting drug sensitivity of a medicine	1. It works on the bagging algorithm used in Ensemble Learning Technique 2. it is a quiet stable algorithm as introduction of new data points may affect one tree, but not all the trees in the Random Forest 3. Missing values are automatically handled and also there is not much affect of noise in the data	Due to the generation of large number of trees to make its prediction, there are two main disadvantages: 1. Computational power and resource complexity increases 2. Training time to train this model is high, also the result depends on majority vote
4	**CNN**	Sarvamangla et al. [12], Bhosale et al. [13], Sumathi et al. [31], Cahyono et al. [48], Dewi et al. [69], Suha et al. [98], Kaur et al. [99], Panicker et al. [101], Hosain et al. [102]	Works on Unstructured dataset	Used primarily in medical image understanding tasks like segmentation, localization, classification and detection	1. They possess efficient image processing due to the presence of filters that have the capability to extract only relevant features 2. They support a very important feature called transfer learning, which means if they are trained on one particular task, that learning can be applied on other similar tasks without additional training	1. Their computational requirements are huge due to the presence of large number of layers and parameters for storing the results and for training and executing 2. The model needs a very large amount of dataset to perform well, since overfitting may occur if the dataset is small and the results with new data may not be accurate

Table 6 (continued)

S.No	Name of the algorithm	References/Authors depicting the use of particular algorithm	Type of Data analyzed by the algorithms	Generalised Application areas(Healt-h sector in particular)	Overall Advantages	Particular Shortcomings
5	**Boosting Algorithms like AdaBoost, Cat-Boost, Gradient Boost, Extreme Gradient Boosting(XGBoost), Stacking**	Danei et al. [6], Khan et al. [24]	Both structured and Unstructured Data	Error reduction in medical data predictions like cardiovascular risks, survival rates of cancer patients	1. They combine the results of multiple weak learners to produce a strong model that can perform with much higher accuracy and computational efficiency 2. They don't require data preprocessing and most languages have built in libraries to implement these routines which can very easily handle missing data as well	2. Outliers may occur sometimes as each model is trying to outperform the result of its predecessor The complexity of the model is high due to the presence of large number of weak learners 3. It is difficult to scale these models

4.6 Research gaps identified and future scope

While carrying out the above SLR, the RID4 and RID6 are answered by recognizing the major gap areas given below that can be further looked up by the research teams to make an improvement in the process of diagnosing and analyzing Polycystic ovaries.

- Most of the studies are considered on a single type of dataset [7]. This gives an insight into considering multi-modality datasets for a better diagnostic approach. Different types of datasets can help the algorithm in predicting PCOS with higher accuracy.
- The different algorithms are compared on similar types of datasets. In future, a single optimization algorithm can be made to analyse different types of datasets.
- The researchers have used the dataset of a particular region or area. The results of such analysis may differ when the algorithms are applied to datasets of some different regions. For this, the researcher can be more attentive towards developing an algorithm that is not specific to data of a particular region but is more focussed towards the parameters considered for evaluation.
- The effect of PCOS on preterm labor/abortion rates is an area that has not yet been answered [41]. This limitation can be a further guiding step for analysts as well as clinicians to detect the problem early and prevent abortions using Machine Learning tools.
- It is found through some of the reviews, that genetics plays an important role in some cases of PCOS [14]. This genetic variance can be studied through ML techniques and can act as an important feature for efficient and accurate detection of PCOS.
- A hybrid approach of feature selection and classification algorithm can be considered, in addition to evaluating the validity of the subsets [35] by the physician.
- A little emphasis is given to identifying those aspects or symptoms that contribute most to the psychological factors [14]. Further research can be carried out that can help in predicting not just PCOS, but also its psychological effects on women's health. This can help healthcare providers to take a cumulative treatment.
- The researchers can extend the study further by studying the relationship between biological issues, anxiety and stress factors and the effect of Vitamin D on PCOS. A new Machine Learning model can be developed, which apart from considering physiological factors can consider these biological aspects also.
- Identification of the multiple cysts using ML can assist clinicians in counseling the patients whether they need assisted reproduction or they need to go for In-vitro fertilization (IVF).

5 Conclusion

The paper focuses on extending a comprehensive view of the various types of classifiers used in detecting cysts, fibroids or follicles in the ovaries. In view of the emerging significance of Machine Learning in the healthcare sector, particularly considering the infertility in women, there is a need of an all-inclusive review and analysis of research in the major cause of it, i.e. Polycystic Ovaries. Keeping this as a motivating factor, this paper tries to reflect upon some useful insights:

- The paper first describes the concept of Machine Learning, including the supervised, unsupervised and deep learning classifiers.

- The existing literature is then systematically examined to answer the research questions (RIDs) that were identified in the onset of this paper. A proper methodology is adopted to perform this SLR.
- The important keywords are then identified that helped in shortlisting the research articles. Following the criteria to include or exclude, some important research papers were selected.
- The selected research papers were critically examined to make a comparison among the key algorithms that are particularly used in healthcare sector and are noted down to find out their advantages and disadvantages.
- Based on this understanding, some major findings were inspected. It was found that the majority of the study revolves around similar types of datasets or feature selection. According to [6], a hybrid model can be constructed that uses both feature selection and classification algorithms to develop a model and also validates the subsets chosen by a healthcare provider.
- D. R. Sarvamangala et al. [25] mentioned that Image captioning can also be included to help and guide physicians to understand the network in a better way.
- The study can be further extended using MRI or CT scan images and using preprocessing techniques to improve the accuracy [27].
- Mathur et al. [34] have suggested automatic longitudinal tracking of follicles during IVF cycles, which can help clinical practitioners in deciding the dosage of gonadotropins given to the patients based on the response of follicles.
- Also, very little insight is provided into those factors that affect women psychologically [36] was concluded by considering the evaluation of the psychological well-being of a woman in the treatment protocol of PCOS.
- There are some other factors mentioned in the future directions section, that can be revisited and re-examined in the future to propose some effective, structured and organized solutions around them.

Data availability Data sharing is not applicable to this article as no datasets were generated or analysed during the current study.

Declarations

Conflicts of interests The authors have no conflict of interest in the publication of this article.

References

1. Rasquin Leon LI, Anastasopoulou C, Mayrin JV. [Online]. https://www.ncbi.nlm.nih.gov/books/NBK459251/#:~:text=Introduction,on%20the%20diagnostic%20criteria%20applied. Accessed 15 Sept 2022
2. Brown S (2021) Machine Learning, Explained. https://mitsloan.mit.edu/ideas-made-to-matter/machine-learning-explained. Accessed Dec 2023
3. Deep Learning vs Machine Learning. [Online]. https://www.ibm.com/topics/deep-learning. Accessed Dec 2022
4. Introduction to Neural Networks [Online]. https://www.ibm.com/topics/neural-networks. Accessed Feb 2023

5. Garg A, Mago V (2021) Role of machine learning in medical research: a survey. https://doi.org/10.1016/j.cosrev.2021.100370
6. Mehr HD, Polat H (2022) Diagnosis of polycystic ovary syndrome through different machine learning and feature selection techniques. Heal Technol 12:137–150. https://doi.org/10.1007/s12553-021-00613-y
7. Tiwari S, Kane L, Koundal D, Jain A, Alhudhaif A, Polat K, Zaguia A, Alenezi F, Althubiti SA (2022) SPOSDS: a smart polycystic ovary syndrome diagnostic system using machine learning. https://doi.org/10.1016/j.eswa.2022.117592
8. Ahmed Z, Kareem MS, Khan HA, Jaskani FH, Saman Z, Mughal B (2022) Detection of uterine fibroids in medical images using deep neural networks. https://doi.org/10.4108/EW201222.31232. https://www.researchgate.net/publication/358264137. Accessed Mar 2023
9. Rathod Y, Komare A, Ajgaonkar R, Chindarkar S, Nagare G, Punjabi N, Karpate Y (2022) Predictive analysis of polycystic ovarian syndrome using CatBoost algorithm. https://ieeexplore.ieee.org/document/9864439. Accessed Dec 2023
10. Zigarelli A, Jia Z, LeeH, (2022) Machine-aided self-diagnostic prediction models for polycystic ovary syndrome: observational study. https://formative.jmir.org/2022/3/e29967. Accessed Apr 2023
11. Nautiyal H, Imam SS, Alshehri S, Ghoneim MM, Afzal M, Alzarea SI, Güven E, Al-Abbasi FA, Kazmi I (2022) Polycystic ovarian syndrome: a complex disease with a genetics approach. Biomedicines 2022(10):540. https://doi.org/10.3390/biomedicines10030540
12. Sarvamangala DR, Kulkarni RV (2022) Convolutional neural networks in medical image understanding: a survey. Evol Intell 15:1–22. https://doi.org/10.1007/s12065-020-00540-3
13. Bhosale S, Joshi L, Shivsharanan A (2022) PCOS (polycystic ovarian syndrome) detection using deep learning. Int Res J of Modernization in Engineering Technology and Science 4:(5)
14. Rajeev D (2022) Psychological implications of poly cystic ovary syndrome: depression, anxiety, stress, body satisfaction, self-esteem, and quality of life. J Adv Res SciSoc Sci (JARSSC) Off Publ Indian Mental Health Res Centre 05(01). ISSN: 2582–2004. https://doi.org/10.46523/jarssc.05.01.11
15. Gopalakrishnan C, Iyapparaja M (2021) Multilevel thresholding based follicle detection and classificationof polycystic ovary syndrome from the ultrasound images using machine learning. Int J Syst Assur Eng Manag. https://doi.org/10.1007/s13198-021-01203-x
16. Yang R, Yu Y (2021) Artificial convolutional neural network in object detection and semantic segmentation for medical imaging analysis.https://doi.org/10.3389/fonc.2021.638182
17. Anaya-Isazaa A, Mera-Jiménez L, Zequera-Diaz M (2021) An overview of deep learning in medical imaging. https://doi.org/10.1016/j.imu.2021.100723
18. Bulsara J, Patel P, Soni A, Acharya S (2021) A review: brief insight into polycystic ovarian syndrome.https://doi.org/10.1016/j.endmts.2021.100085
19. Qayyum A, Qadir J, Bilal M, Al-Fuqaha A (2021) Secure and robust machine learning for healthcare: a survey. https://ieeexplore.ieee.org/document/9153891. Accessed June 2022
20. Chen IY, Pierson E, Rose S, Joshi S, Ferryman K, Ghassemi M (2021) Ethical machine learning in healthcare. https://doi.org/10.1146/annurev-biodatasci-092820-114757
21. Janiesch C, Zschech P, Heinrich K (2021) Machine learning and deep learning. https://doi.org/10.1007/s12525-021-00475
22. Hoeger KM, Dokras A, Piltonen T (2021) Update on PCOS: consequences, challenges, and guiding treatment. J Clin Endocrinol Metab 106(3):e1071–e1083. https://doi.org/10.1210/clinem/dgaa839
23. Zhou SK, Greenspan H, Davatzikos C, Duncan JS, Ginneken BV, Madabhushi A, Prince JL, Rueckert D, Summers RM (2021) A review of deep learning in medical imaging: imaging traits, technology trends, case studies with progress highlights, and future promises. IEEE
24. Inan MSK, Ulfath RE, Alam FI, Bappee FK, Hasan R (2021) Improved sampling and feature selection to support extreme gradient boosting For PCOS diagnosis. IEEE
25. Sarvamangala DR, Kulkarni RV (2021) Convolutional neural networks in medical image understanding: a survey. Springer-Verlag GmbH, DE part of Springer Nature. https://link.springer.com/article/10.1007/s12065-020-00540-3. Accessed Aug 2023
26. Peña AS, Witchel SF, Hoeger KM, Oberfield SE, Vogiatzi MG, Misso M, Garad R, Dabadghao P, Teede H (2020) Adolescent polycystic ovary syndrome according to the international evidence- based guideline. https://doi.org/10.1186/s12916-020-01516-x
27. Boyanapalli MA, Shanthini A (2020) A study of preprocessing techniques and features for ovarian cancer using ultrasound images. Eur J Mol Clin Med 07(10). ISSN 2515–8260
28. Soucie K, Samardzic T, Schramer K, Ly C, Katzman R (2020) The diagnostic experiences of women with polycystic ovary syndrome (PCOS) in Ontario, Canada.https://doi.org/10.1177/1049732320971235

29. Thomas N, Kavitha DA (2020) Prediction of polycystic ovarian syndrome with clinical dataset using a novel hybrid data mining classification technique. Int J Adv Res EngTechnol (IJARET) 11(11). https://doi.org/10.34218/IJARET.11.11.2020.174. http://www.iaeme.com/IJARET/index.asp. Accessed Oct 2023
30. Nusinovici S, Tham YC, Yan MYC, Ting DSW, Li J, Sabanayagam C, Wong TY, Cheng C-Y (2020) Logistic regression was as good as machine learning for predicting major chronic diseases. https://doi.org/10.1016/j.jclinepi.2020.03.002. Elsevier
31. Sumathi M, Chitra P, Prabha RS, Srilatha K (2020) Study and detection of PCOS related diseases using CNN. IOP Conf Ser: Mater Sci Eng 1070(2021):012062. IOP Publishing. https://doi.org/10.1088/1757-899X/1070/1/012062
32. Zhang X, Liang B, Zhang J, Hao X, Xu X, Chang H-M, Leung PCK, Tan J (2020) Molecular and cellular endocrinology, Raman spectroscopy of follicular fluid and plasma with machine-learning algorithms for polycystic ovary syndrome screening. https://doi.org/10.1016/j.mce.2020.111139. Elsevier
33. Elmannai H, El-Rashidy N, Mashal I, Alohali MA, Farag S, El-Sappagh S (2023) Polycystic ovary syndrome detection machine learning model based on optimized feature selection and explainable artificial intelligence. Diagnostics 13:1506. https://doi.org/10.3390/diagnostics13081506
34. Mathur P, Kakwani K, Diplav, Kudavelly S, Ramaraju GA (2020) Deep learning based quantification of ovary and follicles using 3D transvaginal ultrasound in assisted reproduction. IEEE
35. Bharati S, Podder P, Mondal MRH (2020) Diagnosis of polycystic ovary syndrome using machine learning algorithms. IEEE
36 Kodipalli A, Devi S (2021) Prediction of PCOS and mental health using fuzzy inference and SVM. Front Public Health. https://doi.org/10.3389/fpubh.2021.789569
37. Zhang L, Huang J, Liu L (2019) Improved deep learning network based in combination with cost-sensitive learning for early detection of ovarian cancer in color ultrasound detecting system. J Med Syst 43:251. https://doi.org/10.1007/s10916-019-1356-8
38. Harikiran J, Vamsidhar E, Sinivasa Rao B, Saichandana B (2019) Follicle detection in ultrasound images using adaptive clustering algorithms and empirical mode decomposition. Int J Innov Technol Exploring Eng (IJITEE) 9(2)
39. Ramamoorthy S, Vinodhini R, Sivasubramaniam R (2019) Monitoring the growth of polycystic ovary syndrome using mono-modal image registration technique. ACM CoDS- COMAD ’19 6th ACM IKDD CoDS and 24th COMAD, Kolkata, India. https://doi.org/10.1145/3297001.3297024
40. Peiffer-Smadja N, Rawson TM, Ahmad R, Buchard A, Georgiou P, Lescure X, Brigand G, Holmes AH (2019) Machine learning for clinical decision support in infectious diseases: a narrative review of current applications. Clin Microbiol Infect. https://doi.org/10.1016/j.cmi.2019.09.009. (**Elsevier**)
41. Denny A, Raj A, Ashok A, Ram CM, George R (2019) i-HOPE: detection and prediction system for polycystic ovary syndrome (PCOS) using machine learning techniques. IEEE
42. Omar KS, Mondal P, Khan NS, Rizvi MRK, Islam MN (2019) A machine learning approach to predict autism spectrum disorder. IEEE
43. Toosy S, Sodi R, Pappachan JM (2018) Lean polycystic ovary syndrome (PCOS): an evidence-based practical approach. J Diabetes Metab Disord. https://doi.org/10.1007/s40200-018-0371-5
44. Shailaja K, Seetharamulu B, Jabbar MA (2018) Machine learning in healthcare: a review. IEEE
45. Callahan A, Shah NH (2017) Machine learning in healthcare.https://doi.org/10.1016/B978-0-12-809523-2.00019-4
46 Costea OD, Ivanoschi A, Badiu D, Băjan S (2017) Surgical management of ovarian cysts in post-menopausal women. Gineco.eu 13:165–168. https://doi.org/10.18643/gieu.2017.165. (@ **2017 Romanian Society of Ultrasonography in Obstetrics and Gynecology**)
47. Wisesty UN, Nasri J, Adiwijaya (2017) Modified backpropagation algorithm for polycystic ovary syndrome detection based on ultrasound images. © Springer International Publishing AG 2017 Herawan T et al. (eds), Recent advances on soft computing and data mining, advances in intelligent systems and computing, 549. https://doi.org/10.1007/978-3-319-51281-5_15
48. Cahyono B, Adiwijaya, Mubarok MS, Wisesty UN (2017) An implementation of convolutional neural network on PCO classification based on ultrasound image. 2017 Fifth international conference on information and communication technology (ICoICT). IEEE. https://doi.org/10.1109/ICoICT.2017.8074702
49. Ravishankar H, Sudhakar P, Venkataramani R, Thiruvenkadam S, Annangi P, Babu N, Vaidya V (2016) Understanding the mechanisms of deep transfer learning for medical images. Springer international publishing AG 2016 Carneiro G et al. (Eds): LABELS 2016/DLMIA 2016, LNCS 10008, pp 188–196. https://doi.org/10.1007/978-3-319-46976-8

50. Shan B, Cai J-h, Yang S-Y, Li Z-R (2015) Risk factors of polycystic ovarian syndrome among Li People. Elsevier. https://doi.org/10.1016/j.apjtm.2015.07.001
51. Setiawati E, Adiwijaya, Tjokorda ABW (2015) Particle swarm optimization on follicles segmentation to support PCOS detection. IEEE
52 Padmapriya B, Kesavamurthy T (2015) Diagnostic tool for PCOS classification. Springer International Publishing Switzerland
53. Witchel SF, Oberfield SE, Pena AS (2019) Polycystic ovary syndrome: pathophysiology presentation, and treatment with emphasis on adolescent girls. J Endocr Soc. https://doi.org/10.1210/js.2019-00078
54. Zhu T, Cui J, Goodarzi MO (2020) Polycystic ovary syndrome and risk of type 2 diabetes, coronary heart disease, and stroke. DivEndocrinol Diab Metab. https://doi.org/10.2337/db20-0800
55. de Matos J, Ataky STM, de Souza Britto A Jr, de Oliveira LES, Koerich AL (2021) Machine learning methods for histopathological image analysis: review. MDPI Electron 10:562. https://doi.org/10.3390/electronics10050562
56. Simons PIHG, Cornelissen MEB, Olivier Valkenburg N, Onland-Moret C, van der Schouw YT, Stehouwer CDA, Burgess S, Brouwers MCGJ (2021) Causal relationship between polycystic ovary syndrome and coronary artery disease: a Mendelian randomisation study. Clin Endocrinol J. https://doi.org/10.1111/cen.14593
57. Prasanth S, Thanka RM, Edwin EB, Ebenezer V (2021) Prognostication of diabetes diagnosis based on different machine learning classification algorithms. Ann RSCB 25(5):372–395. ISSN:1583–6258
58. Azziz R (2006) Diagnosis of polycystic ovarian syndrome: the Rotterdam criteria are premature. J Clin Endocrinol Metab 91(3):781–785. https://doi.org/10.1210/jc.2005-2153
59 Hart R, Doherty DA (2015) The potential implications of a PCOS diagnosis on a woman's long-term health using data linkage. J Clin Endocrinol Metab. https://doi.org/10.1210/jc.2014-3886
60. Adiwijaya, Purnama B, Hasyim A, Septiani MD, Wisesty UN, Astuti W (2015) Follicle detection on the USG images to support determination of polycystic ovary syndrome. J Phys: Conf Ser 622. https://doi.org/10.1088/1742-6596/622/1/012027
61. Sayyah-Melli M, Alizadeh M, Pourafkary N, Ouladsahebmadarek E, Jafari-Shobeiri M, Abbassi J, Kazemi-Shishvan M, Sedaghat K (2015) Psychosocial factors associated with polycystic ovary syndrome: a case control study. J Caring Sci 4(3): 225–231https://doi.org/10.15171/jcs.2015.023. http://journals.tbzmed.ac.ir/JCS
62 Jolliffe IT, Cadima J (2016) Principal component analysis: a review and recent developments. Phil Trans R Soc A 374:20150202. https://doi.org/10.1098/rsta.2015.0202
63. Gibson-Helm M, Teede H, Dunaif A, Dokras A Delayed diagnosis and a lack of information associated with dissatisfaction in women with polycystic ovary syndrome. J Clin Endocrinol Metab Endocr Soc. https://doi.org/10.1210/jc.2016-2963
64. Rihana S, Moussallem H, Skaf C, Yaacoub C (2013) Automated algorithm for ovarian cysts detection in ultrasonogram. 2013 2nd international conference on advances in biomedical engineering. IEEE. https://doi.org/10.1109/ICABME.2013.6648887
65 Acharya UR, Molinari F, Sree SV, Swapna G, Saba L, Guerriero S, Suri JS (2015) Ovarian tissue characterization in ultrasound: a review. Technol Cancer Res Treat 14(3):251–261. https://doi.org/10.1177/1533034614547445
66. Chaudhary K, Poirion OB, Lu L, Garmire LX (2017) Deep Learning based multi-omics integration robustly predicts survival in liver cancer. http://clincancerres.aacrjournals.org/. https://doi.org/10.1158/1078-0432.CCR-17-0853
67. Engmann L, Jin S, Sun F, Legro RS, Polotsky AJ, Hansen KR, Coutifaris C, Diamond MP, Eisenberg E, Zhang H, Santoro N (2017) Racial and ethnic differences in the polycystic ovary syndrome (PCOS) metabolic phenotype. Am J Obstet Gynecol. https://doi.org/10.1016/j.ajog.2017.01.003
68. Mohammad MB, Seghinsara AM Polycystic ovary syndrome (PCOS), diagnostic criteria, and AMH. Asian Pac J Cancer Prev 18.https://doi.org/10.22034/APJCP.2017.18.1.17
69. Dewi RM, Adiwijaya, Wisesty UN, Jondri (2018) Classification of polycystic ovary based on ultrasound images using competitive neural network. IOP Conf Ser J Phys Conf Ser 971:012005. https://doi.org/10.1088/1742-6596/971/1/012005
70. Balogun JA, Egejuru NC, Idowu PA (2018) Comparative analysis of predictive models for the likelihood of infertility in women using supervised machine learning techniques. Comput Rev J 2. https://purkh.com/index.php/tocomp. Accessed July 2023
71. Teede HJ, Misso ML, Costello MF, Dokras A, Laven J, Moran L, Piltonen T, Norman RJ (2018) Recommendations from the international evidence-based guideline for the assessment and management of polycystic ovary syndrome. Human Reprod 33(9):1602–1618. https://doi.org/10.1093/humrep/dey256. https://academic.oup.com/humrep/article/33/9/1602/5056069. Accessed Nov 2023

72. Wolf WM, Wattick RA, Kinkade ON, Olfert MD (2018) Geographical prevalence of polycystic ovary syndrome as determined by region and race/ethnicity. Int J Environ Res Public Health 15:2589. https://doi.org/10.3390/ijerph15112589
73. Andreotti RF, Timmerman D, Strachowski LM, Froyman W, Benacerraf BR, Bennett ML, Bourne T, Brown DL, Coleman BG, Frates MC, Goldstein SR, Hamper UM, Horrow MM, Hernanz-Schulman M, Reinhold C, Rose SL, Whitcomb BP, Wolfman WL, Glanc Phyllis (2020) O-RADS US risk stratification and management system: a consensus guideline from the ACR ovarian-adnexal reporting and data system committee. Radiology 294:168–185. https://doi.org/10.1148/radiol.2019191150
74 Ji HR, Woo HL, Park YJ, Hwang DS, Lii JM, Lee CH, Jang JB, Park KS (2018) Characteristics of heart rate variability in women with polycystic ovary syndrome. A retrospective cross sectional study. Medicine 97(38):e12510. https://doi.org/10.1097/MD.0000000000012510
75 Sarker IH (2021) Machine learning: algorithms, real-world applications and research directions. SN Comput Sci 2:160. https://doi.org/10.1007/s42979-021-00592-x
76. Sreejith S, Nehemiah HK, Kannan A (2022) A clinical decision support system for polycystic ovary syndrome using red deer algorithm and random forest classifier. Elseiver. https://doi.org/10.1016/j.health.2022.100102
77. Barber TM, Franks S (2021) Obesity and polycystic ovary syndrome. Clin Endocrinol 95:531–541. https://doi.org/10.1111/cen.14421
78 Maadi M, Khorshidi HA, Aickelin U (2021) A review on human–AI interaction in machine learning and insights for medical applications. Int J Environ Res Public Health 18:2121. https://doi.org/10.3390/ijerph18042121
79 Saber A, Sakr M, Abo-Seida OM, Keshk A, Chen H (2021) A novel deep-learning model for automatic detection and classification of breast cancer using the transfer-learning technique. Dig Object Identifier. https://doi.org/10.1109/ACCESS.2021.3079204
80. Sendur SN, Yildiz BO (2020) Influence of the ethnicity on the different aspects of the polycystic ovary syndrome: a systematic review. Reprod BioMed J. https://doi.org/10.1016/j.rbmo.2020.12.006
81. Lupean R-A, Stefan P-A, Oancea MD, Mălut AM, Lebovici A, Puscas ME, Mihu Csutak C, CM, (2020) Computer tomography in the diagnosis of ovarian cysts: the role of fluid attenuation values. Healthcare 8:398. https://doi.org/10.3390/healthcare8040398
82. Dapas M, Lin FTJ, Nadkarni GN, Sisk R, Legro RS, Urbanek MM, Hayes G, Dunaif A (2020) Distinct subtypes of polycystic ovary syndrome with novel genetic associations: An unsupervised, phenotypic clustering analysis. J PLOS Med. https://doi.org/10.1371/journal.pmed.1003132
83. Jusman Y, Indra Z, Salambue R, MohdKanafiah SNA, Nurkholid MFA (2020) Comparison of multi layered percepton and radial basis function classification performance of lung cancer data. IOP Conf Ser J Phys Conf Ser 1471:012043. https://doi.org/10.1088/1742-6596/1471/1/012043
84. Ricciardi C, Valente AS, Cantoni KEdV, Green R, Fiorillo A, Picone I, Santini S, Cesarelli M (2020) Linear discriminant analysis and principal component analysis to predict coronary artery disease. Health Inform J 26(3):2181–2192. https://doi.org/10.1177/1460458219899210
85. Munjal A, Khandia R, Gautam B (2020) A machine learning approach for selection of polycystic ovarian syndrome (pcos) attributes and comparing different classifier performance with the help of weka and pycaret. 2277-8179. https://doi.org/10.36106/ijsr/5416514
86. Prapty AS, Shitu TT (2020) An efficient decision tree establishment and performance analysis with different machine learning approaches on polycystic ovary syndrome. IEEE. https://doi.org/10.1109/ICCIT51783.2020.9392666
87 Wang D, Li R, Wang J, Jiang Q, Gao C, Yang J, Ge L, Qing H (2020) Correlation analysis between disease severity and clinical and biochemical characteristics of 143 cases of COVID-19 in Wuhan, China: a descriptive study. BMC Infect Dis 20:519. https://doi.org/10.1186/s12879-020-05242-w
88. Janssens ACJ, Martens FK (2020) Reflection on modern methods: Revisiting the area under the ROC Curve. Int J Epidemiol 1397–1403. https://doi.org/10.1093/ije/dyz274
89. Nandipati SCR, XinYing C, Wah KK (2020) Polycystic ovarian syndrome (PCOS) classification and feature selection by machine learning techniques. Appl Math Comput Intell 9:65–74
90. Álvarez JD, Matias-Guiu JA, Cabrera-Martín MN, Risco-Martín JL, Ayala JL (2019) An application of machine learning with feature selection to improve diagnosis and classification of neurodegenerative disorders. BMC Bioinformatics 20:491. https://doi.org/10.1186/s12859-019-3027-7
91. Ul Haq A, Li JP, Memon MH, Nazir S, Sun R (2018) A hybrid intelligent system framework for the prediction of heart disease using machine learning algorithms. Mob Inform Syst 2018. https://doi.org/10.1155/2018/3860146lgorithms
92. Escobar-Morreale HF (2018) Polycystic ovary syndrome: definition, aetiology, diagnosis and treatment. Nat Rev Endocrinol 14. https://doi.org/10.1038/nrendo.2018.24

93. Prajapati D (2022) Development and evaluation of a polyherbal tablet for polycystic ovarian syndrome (PCOS). J Nat Remedies 22(2). http://www.informaticsjournals.com/index.php/jnr. Accessed Dec 2023
94 Armanini D, Boscaro M, Bordin L, Sabbadin C (2022) Controversies in the pathogenesis, diagnosis and treatment of PCOS: focus on insulin resistance, inflammation, and hyperandrogenism. Int J Mol Sci 23:4110. https://doi.org/10.3390/ijms23084110
95. Armstrong J, Cortes C, Hawkins K, Younis A (2022) Serum anti-Mullerian Hormone (AMH) levels are effective in predicting the diagnosis of four polycystic ovarian syndrome (PCOS) phenotypes. Gynecol Reprod Health 6(5):1–5
96 Nssibi M, Manita G, Korbaa O (2023) Advances in nature-inspired metaheuristic optimization for feature selection problem: a comprehensive survey. Comput Sci Rev. https://doi.org/10.1016/j.cosrev.2023.100559
97. Rađenović S, Pupovac M, Andjić M, Bila J, Srećković S, Gudović A, Dragaš B, Radunović N (2022) Prevalence, risk factors, and pathophysiology of nonalcoholic fatty liver disease (NAFLD) in women with polycystic ovary syndrome (PCOS). Biomedicines 10:131. https://doi.org/10.3390/biomedicines10010131
98. Suha SA, Islam MN (2022) An extended machine learning technique for polycystic ovary syndrome detection using ovary ultrasound image. Sci Rep 12:17123. https://doi.org/10.1038/s41598-022-21724-0
99. Kaur R, Kumar R, Gupta M (2022) Food Image-based diet recommendation framework to overcome PCOS problem in women using deep convolutional neural network. Elseiver. https://doi.org/10.1016/j.compeleceng.2022.108298
100. Nasim S, Almutairi MS, Munir K, Raza A, Younas F (2022) A novel approach for polycystic ovary syndrome prediction using machine learning in bioinformatics. IEEE Access J. https://doi.org/10.1109/ACCESS.2022.3205587
101. Panicker PH, Shah K, Chandani SK CNN based image descriptor for polycystic ovarian morphology from transvaginal ultrasound. https://doi.org/10.1109/CSCITA55725.2023.10104931
102. Hosain AKMD, Mehedi MHD, Kabir II PCONet: a convolutional neural network architecture to detect polycystic ovary syndrome (PCOS) from ovarian ultrasoundimages. https://doi.org/10.1109/ICEET56468.2022.10007353
103. Kumari R, Singh J, Kumari A (2023) SmS: SMOTE-stacked hybrid model for diagnosis of polycystic ovary syndrome using feature selection method. Elseiver. https://doi.org/10.1016/j.eswa.2023.120102
104. Harshvardhan GM, Gourisaria MK, Pandey M, Rautaray SS (2020) A comprehensive survey and analysis of generative models in machine learning. Comput Sci Rev. https://doi.org/10.1016/j.cosrev.2020.100285
105. Smiti A (2020) When machine learning meets medical world: current status and future challenges. Comput Sci Rev. https://doi.org/10.1016/j.cosrev.2020.100280
106. Alimirzaei F, Kieslich CA (2023) Machine learning models for predicting membranolytic anticancer peptides. Comput Aided Chem Eng. https://doi.org/10.1016/B978-0-443-15274-0.50428-5
107. Rastegar-Kashkouli A, Jafari M, Karami S, Yousefi P, Taravati AM, Khavaran A, Rastegar D, Jafari MR, Alemohammad SY, Abbasabad GD, Shahbaz M, Kalan ME Association between type 2 diabetes mellitus and multiple myeloma: fact or fiction. J Nephropharmacol. https://doi.org/10.34172/npj.2023.10604
108. Maydanchi M et al (2023) "Comparative study of decision tree, AdaBoost, Random Forest, Naïve Bayes, KNN, and Perceptron for Heart Disease Prediction" SoutheastCon 2023, Orlando, FL, USA, pp 204–208. https://doi.org/10.1109/SoutheastCon51012.2023.10115189.
109. Fouladi S, Safaei AA, Arshad NI et al (2022) The use of artificial neural networks to diagnose Alzheimer's disease from brain images. Multimed Tools Appl 81:37681–37721. https://doi.org/10.1007/s11042-022-13506-7