IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 13 July 2023, accepted 27 July 2023, date of publication 11 August 2023, date of current version 17 August 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3304536

TOPICAL REVIEW

A Review on the Detection Techniques of Polycystic Ovary Syndrome Using Machine Learning

SAMIA AHMED[1,*], MD. SAZZADUR RAHMAN[2,*], ISMATE JAHAN[1], M. SHAMIM KAISER[2], (Senior Member, IEEE), A. S. M. SANWAR HOSEN[3], (Member, IEEE), DEEPAK GHIMIRE[4], AND SEONG-HEUM KIM[4]

[1]Department of Information and Communication Technology (ICT), Bangladesh University of Professionals, Dhaka 1216, Bangladesh
[2]Institute of Information Technology, Jahangirnagar University, Savar, Dhaka 1342, Bangladesh
[3]Department of Artificial Intelligence and Big Data, Woosong University, Daejeon 34606, South Korea
[4]School of AI Convergence, College of Information Technology, Soongsil University, Seoul 06978, South Korea

Corresponding authors: Md. Sazzadur Rahman (sazzad@juniv.edu) and Seong-Heum Kim (seongheum@ssu.ac.kr)

This work was supported in part by the National Research Foundation of Korea (NRF) grant funded by the Korean Government (MSIT) under Grant 2021R1G1A1009828 and Grant 2021R1A4A1032252; and in part by the Woosong University Academic Research Fund, South Korea.

*Samia Ahmed and Md. Sazzadur Rahman contributed equally to this work.

ABSTRACT Polycystic Ovary Syndrome (PCOS) is a critical hormonal disorder of women that significantly impacts life. In this new generation, women are more prone to PCOS. It is the cause of various problems, including infertility. Early detection of PCOS can reduce complexity. Therefore, an early and proper PCOS detection system is essential to minimize complications. Among all the detection techniques Machine Learning (ML) has an excellent performance in detection for its feature extraction capability. Therefore, considerable research has been carried out to detect PCOS using ML. Various ML approaches like Convolutional Neural Network, Support Vector Machine, K-Nearest-Neighbors, Random Forest, Logistic Regression, Decision Tree, Naive Bayes, etc., are used in detecting PCOS. This research aims to call attention to the researchers by presenting a descriptive and contextual overview of all the existing technologies on PCOS detection by ML algorithms. A comprehensive analysis is carried out of how various ML approaches have been used in PCOS detection over the last few decades, and the techniques are discussed thoroughly. A complete examination was studied on different datasets used in PCOS detection. The performance of several algorithms is compared in quantitative and qualitative approaches. Finally, the significant difficulties and future research scopes are discussed to draw a conclusion.

INDEX TERMS Polycystic ovary syndrome, machine learning, convolutional neural network, women's health.

I. INTRODUCTION

Polycystic Ovary Syndrome (PCOS) is one of the most common hormonal disorders in women. It is an endocrine and metabolic disorder in premenopausal women [1]. It is tough to recognize and treat PCOS. The correct cause of PCOS is not clear yet, but there is a close association with family history and hereditary qualities, hormones that are expanded during our advancement within the womb sometime recently birth, and way of life or environment [2]. It is commonly related to raised levels of two male hormones within the body. This hormone causes their body to skip menstrual periods and makes it harder for them to urge pregnant. Women with PCOS create a higher-than-normal number of male hormones. The side effects of PCOS are shown in numerous distinctive ways. A few women will have minor or gentle side effects, while others will have several serious side effects. Mood changes, sadness, uneasiness, and low self-esteem is some of the side effects of PCOS. Indications can, moreover, alter at distinctive stages of a woman's life. Some of the primary symptoms

The associate editor coordinating the review of this manuscript and approving it for publication was Yiming Tang.

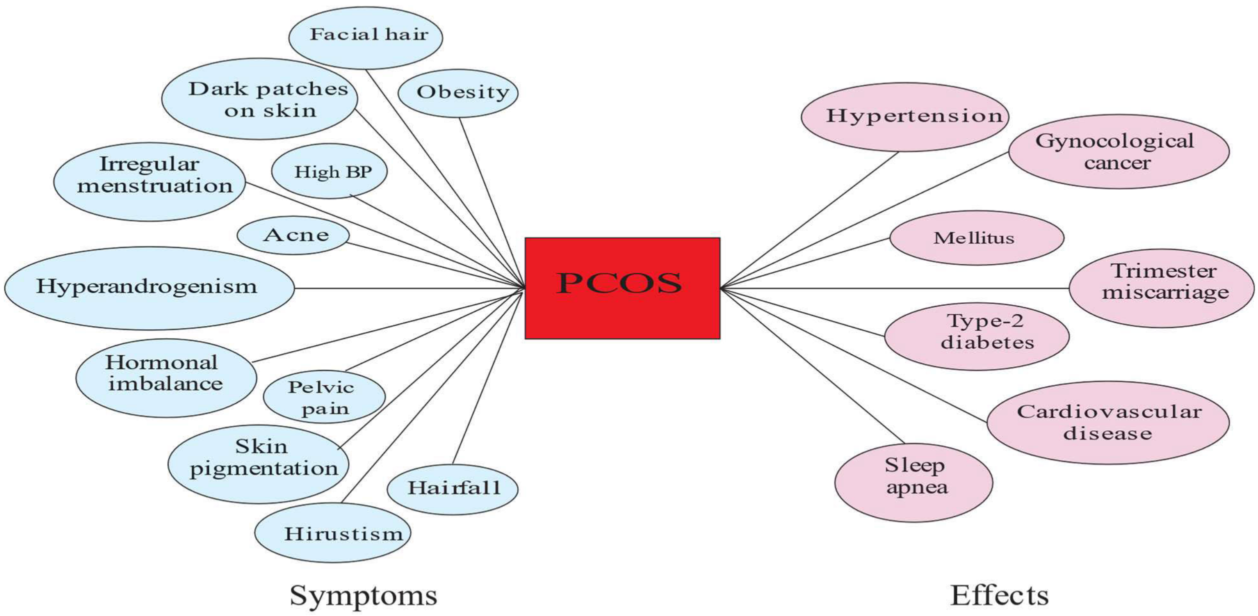

FIGURE 1. Symptoms and some basic effects that are found in women affected with PCOS. The left elements of PCOS are symptoms, and the right ones are effects of PCOS.

of PCOS are no period or delayed period, immature ovarian eggs that do not ovulate, different 'cysts' on the ovaries, trouble in getting pregnant, excessive hair development often on the upper lip, chest, back, or buttocks, weight gain, thinning hair and hair misfortune from the head, oily skin or acne are some of the symptoms [1]. Obesity, skin darkening, and skin pigmentation are more symptoms of PCOS. The effect of PCOS includes type-2 diabetes, cardiovascular disease, sleep apnea, mellitus, and trimester miscarriage. Fig. 1 visually represents the symptoms and effects of PCOS. By focusing on this figure, researchers can easily identify the symptoms and the effects caused by PCOS.

The primary need is to distinguish between PCOS and non-PCOS and treat- PCOS as early as possible. It requires a few tests and imaging methods as conceivable since the circumstance caused ovary disorder, which increases the chance of pregnancy complications, obstetric tumors, and mental trouble. Although much research was conducted to analyze PCOS utilizing different ML calculations, there is still a need for change in terms of exactness and precision based on medical information [3]. Most women are unaware of their regenerative organs and the issues related to them. It causes infertility, uterus tumors, and closes with cancer. Parcels of therapeutic tests and time can discourage a PCOS-influenced woman from curing herself legitimately [4]. Fig. 2 represents the rate of PCOS affecting people around the world. It is seen that Asian women are infected most with PCOS, the rate is almost 31.3%, and White Americans are less infected with 4.8%. African Americans are infected at 6.8%, and Spanish women at 6.8%. This figure identifies race and genetic as two major elements behind PCOS in women.

The authors considered all the published papers from 2015 to 2022 related to using ML for detecting and predicting

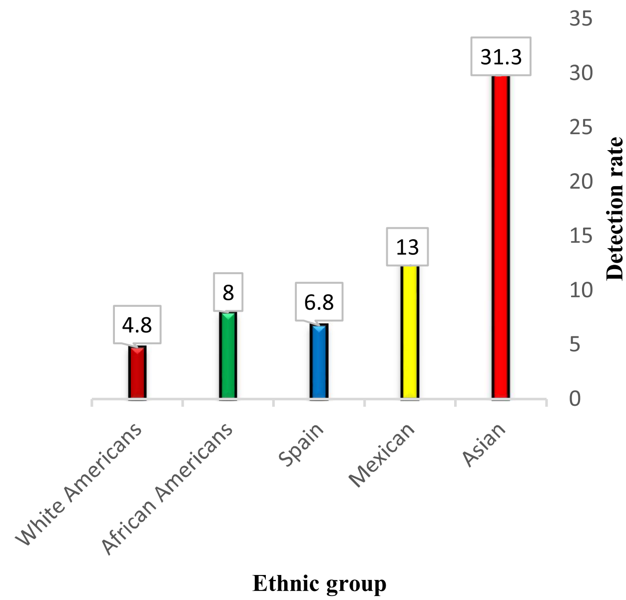

FIGURE 2. PCOS detection rate among women from different ethnic groups around the world.

PCOS. However, there are no such review articles regarding PCOS detection using ML that uses statistical parameters for comparing the performances of various kinds of literature. To the best of the authors' knowledge, there are a few numbers of review papers on PCOS detection. Some directly use ML, some use computer-assisted technology, and some use Artificial Intelligence (AI) and image processing. Gotte and Jyothi [5] surveyed the topic of PCOS detection segmentation and classifications utilizing divergent ML executions. But the survey does not include any potential comparison of the existing literature. Vaswani et al. [6] reviewed papers that were published between 2020 to 2021. The research provides a review of the papers that use ML techniques for the diagnosis of PCOS. They considered the papers that use Deep Learning (DL) and ML for PCOS detection. They initiated the idea of using object detection in the field of detecting PCOS. Pulluparambil and Bhat [7] evaluated research works that are related to image processing. They thoroughly analyzed the existing detection techniques, phases of image processing, and existing algorithms. They found some research gaps and proposed an architecture for PCOS detection.

Rabiu et al. [8] analyzed the research articles that used computer-assisted techniques for the detection of follicles and PCOS diagnoses in the ultrasound images of the ovary. They considered papers ranging from the year 1997 to 2014. They completed their research by overviewing the existing techniques. They made five subsections of the techniques and discussed the process. They focused on the detection rate of follicles but did not show any performance matrices. Agrawal et al. [9] revised the methods of AI for the detection of PCOS. They showed the performances but did not show any comparison between the performances. The existing review papers have significant shortcomings. Firstly, they reviewed a handful of papers and only discussed those works but did not discuss any systematic information, like the shortcomings and comparison of performances. Secondly, they did not show any quantitative or qualitative analysis of the techniques. Thirdly, they did neither provide a detailed discussion about the dataset nor a proper outline for future research scope.

This research chose a total of 34 papers to review, published in the time range between 2003 to 2023. Use of ML, for PCOS detection is increasing day by day. Fig. 3 visualizes the chronological representation of articles that are selected for review. It is seen that from the years 2021 and 2023, a maximum of 8 papers are chosen followed by 2020. For this work, journal and conference papers related to the application of ML in PCOS detection were searched in renowned search engines like Google Scholar, IEEE Explorer, and Science Direct, and 60 research articles were collected. After removing duplicates, 56 articles were there for review. After reviewing the abstract, 17 articles were excluded as they were not related to PCOS detection using ML. Then 39 research articles were assessed, and five articles were found as review articles. After completing all the selection procedures, 34 research articles were finally selected for review. We thoroughly examined the 25 papers and found the research gaps. The whole selection process is described in fig. 4. This figure gives a clear idea how researchers can find related works and further select their final related works for their research. The first two selection processes are the identification phase, followed by the screening phase, and at last, the final selection phase.

After selecting the 34 papers, the next step was to find the research gaps. Then the papers were categorized based on the used algorithms like Convolutional Neural Network (CNN) or Artificial Neural Network (ANN) based articles. Then the

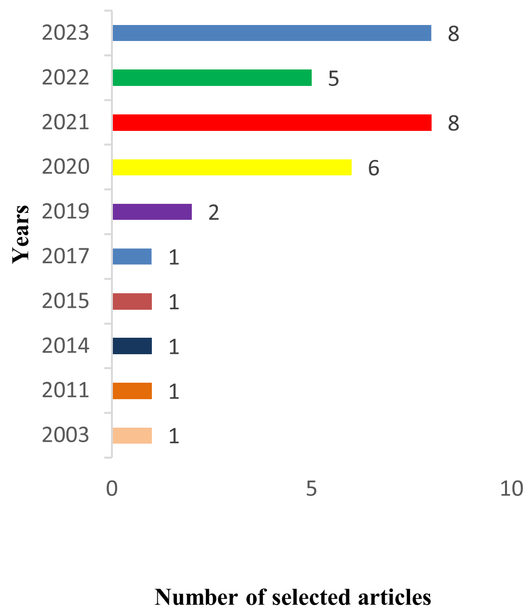

FIGURE 3. Chronological diagram of articles that are considered for this research from 2003 to 2023.

papers that used the same type of technologies were merged. They were placed in the same paragraphs and linked based on the technologies. Next, we evaluated the algorithms used by using charts and tables. We conducted both quantitative and qualitative analyses. Then the research challenges were identified, and possible solutions were suggested.

Fig. 5 presents the detailed research process of this work. Readers will get a vivid concept of how this work has been completed.

This study concentrates on works regarding PCOS detection and offers a descriptive method. One can comprehend the algorithm for detecting PCOS as well as the benefits and drawbacks of the current approaches. Depending on their application, researchers can select the optimal algorithm from the available PCOS detection methods. This research also offers an accurate description of the datasets relevant to the topic and illustrates the range and unrecognized problems in the area. This study incorporates all the work on PCOS detection algorithms conducted previously. The major contributions of the study are discussed below:

- A precise concept of PCOS detection techniques that use traditional methods is analyzed.
- Existing algorithms in the context of ML are explored.
- A description of different types of datasets is given.
- The performance of the dataset in various ML algorithms is found.
- A comparison of performances of the existing PCOS detection techniques has been highlighted.
- Challenges related to PCOS detection using ML are focused, and research ideas are provided for future work.

The rest of the paper has been organized as follows. Section II reports on the existing approaches to PCOS detection. Section III presents an elaborate discussion regarding

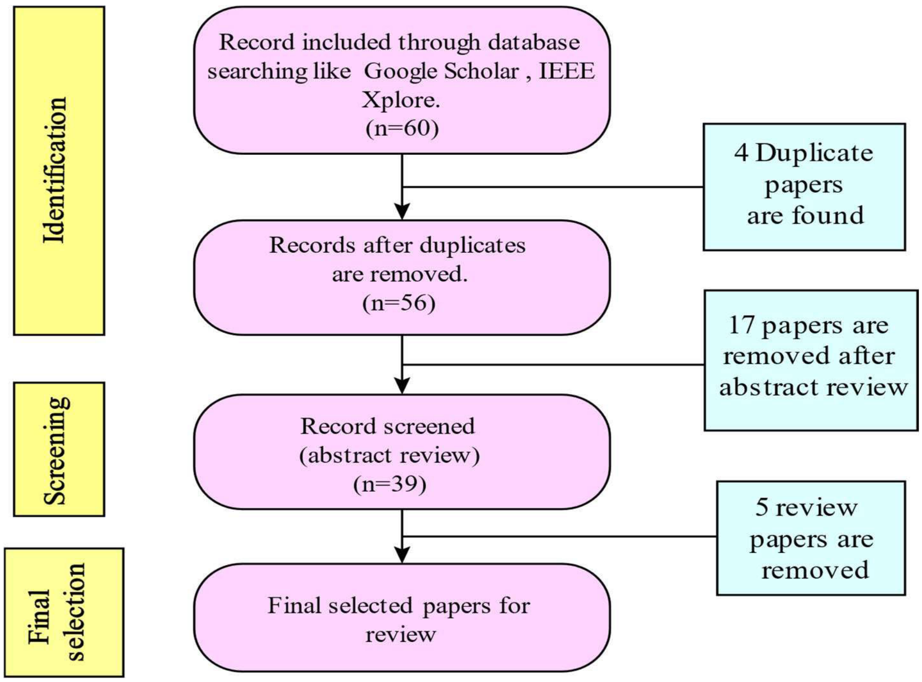

FIGURE 4. Visual representation of the paper selection process for this research work.

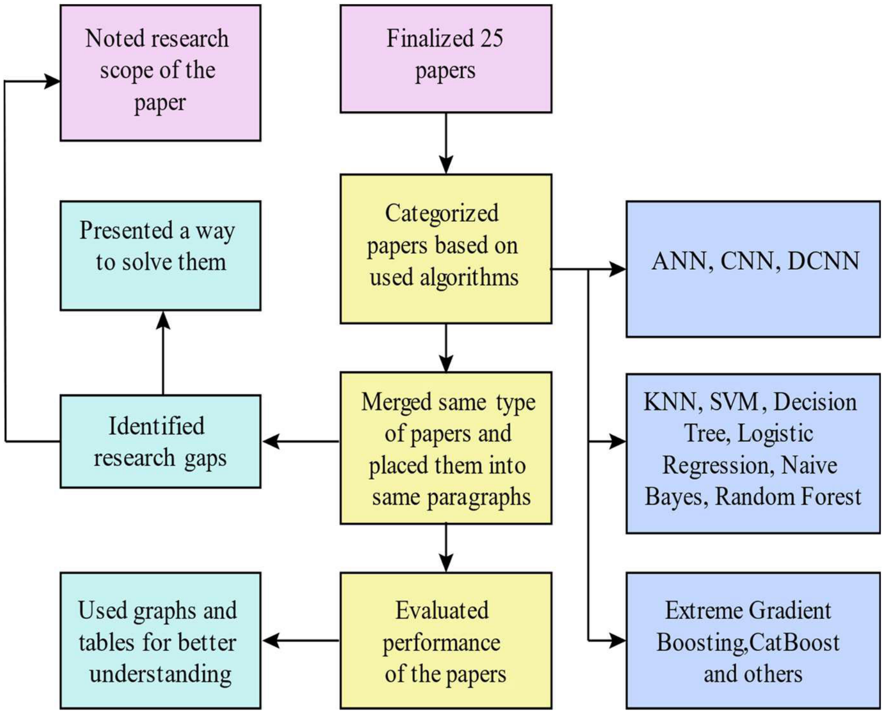

FIGURE 5. A flowchart of the research procedure of this work. The arrow identifies the continuity of the work.

PCOS detection techniques. Section IV is organized by mentioning effective datasets, and their descriptions. Section V summarizes the performance analysis of different approaches that are discussed in section II. Section VI identifies the challenges and probable solutions in implementing these algorithms and future directions are given for researchers. Finally, section VII concludes the study.

II. LITERATURE REVIEW

In this section, the existing literature is discussed along with the used approaches and algorithms. Also, descriptive comparison and table are included to give a brief analysis of literature. This section includes a table that summarizes all the existing works and gives remarks about the work.

Sumathi et al. [10] proposed CNN-based image processing feature extraction for PCOS. Here, the model is based on image processing features extraction to determine a cyst or not. For feature extraction watershed algorithm is used, and for measuring parameters, OpenCV is used. The model can only detect cysts but cannot classify whether PCOS is or not. The model shows 85% accuracy. Cahyono et al. [11] suggested an approach in which the characteristic extraction is executed mechanically using CNN having an f1-rating of 100% and a regular of 76.36% on a 5-overlap cross-validation. The dataset is composed of a total of 40 non-PCO information and 14 polycystic ovaries (PCO) information. Dewi et al. [12] outlined a system to classify PCO by including the Gabor Wavelet strategy (GWS) for feature extraction and CNN. Firstly, the images are preprocessed by using various techniques that include data cleaning, image binarization, morphology, inverted image, grayscaling, etc. Secondly, segmentation is performed. Thirdly, GWS is optimized for feature extraction, and lastly, CNN is applied. CNN obtains the highest accuracy, 80.84%, and the time handle is 60.64 seconds. Hossain et al. [13] created a CNN-based model called PCONet to detect PCO from ovarian ultrasound pictures. Additionally, they improved InceptionV3, a 45-layer pre-trained CNN, by classifying ultrasound images of polycystic ovarian tissue using the transfer learning approach. In the comparison of these two models based on quantitative performance evaluation metrics, they found that PCONet outperformed them both with an accuracy of 98.12% on test images, while InceptionV3 underwent fine-tuning to achieve an accuracy of 96.56%. Bhosale et al. [14] suggested PCOS detection using an LR classifier. This method is proposed for the automated quality evaluation of PCOS data using a Deep Convolutional Neural Network (DCCN). They recommended a deep learning model that, when used with the Inception model, could identify the PCOM from the ultrasound image with an accuracy of 84.81%. Alamoudi et al. [15] suggested a fusion model that combines the ultrasound image with clinical information to determine whether the patient has PCOS or not. The best model created, which combined clinical features with picture feature extraction using MobileNet architecture, obtained 82.46% accuracy. Chitra et al. [16] implemented transfer learning methods such as Alexnet, Inception V3, Resnet50, VGG16, and Hybrid Models to offer a method of PCOS prediction. Here, an attempt is made to present an approach that incorporates hybrid models in order to train models and increase accuracy. The feature extraction technique's outcomes are found by utilizing the performance coefficients. The accuracy of the test data set is 93%.

Krishnaveni et al. [17] worked for a PCOS Screener based on the ANN, a classification model with the support of a questionnaire-based diagnostic tool. The performance of the model was compared with the Bayesian classifier, where the ANN model performed better. Thara [18] proposed a system for classifying PCOS using feature extraction by Structural Normalized Square Similarity Detection Approach (SNSSDA) and ANN. Firstly, image preprocessing, like scaling and noise reduction, is performed. Following that, the segmentation approach is used. Adaptive k-means clustering techniques were used to find follicles in ultrasound pictures. Following segmentation, SNSSDA-based feature extraction and an ANN-based classifier are utilized to detect follicles in the ovary picture.

Maheshwari et al. [19] showed the Naïve Bayes (NB) classifier and ANN in PCOS detection. However, utilizing this, we can determine the inspiration and negativity of the PCOS in earlier arrangements based on the measurement. Precision is about 98.63%, accuracy and specificity are 100%, F- measure is 68.76%, and review is 55%. Hassan and Mirza [20] suggested an unused method, such as XGBRF and CatBoost model, for early identification of PCOS. Here ML approaches for example Support Vector Machine (SVM), Classification and Regression Trees (CART), NB classification, Random Forest (RF), and Logistic Regression (LR) to diagnose PCOS based on the clinical information of patients. RF shows a maximum of 96% accuracy. Satish et al. [21] reported the PCOS detection process by using seven classification algorithms, such as KNN, SVM, RF, Gaussian NB, and Neural Network (NN) and five feature selection methods. Rapid Miner and Python (Spyder as IDE) - Scikit Learn package were used as the tool. The first one was the most accurate, with 93.12% accuracy on the entire dataset.

A technique that can help in the early identification and forecasting of PCOS therapy using an ideal and minimal set of criteria is proposed by Thakre et al. [22]. They propose a technique that can help in the early identification and forecasting of PCOS using an ideal and minimal set of criteria. Five distinct ML classifiers, including RF, SVM, LR, Gaussian NB, and K-Nearest-Neighbor (KNN), have been used to determine whether a woman has PCOS. Using the Chi-Square technique, the top 30 features from the dataset's 41 features were determined. Among four others, the RF classifier was judged to be the most trustworthy and accurate, with an accuracy rate of 90.9%. Purnama et al. [23] recommended PCOS detection by using ultrasound images based on follicle detection. Here several classification methods, such as SVM and KNN, have been applied. Among them, the best accuracy gained from SVM - RBF kernel - approximately 82.55%. Dutta et al. [24] planned to employ Synthetic Minority Oversampling Technique (SMOTE) and five ML techniques, namely LR, RF, Decision Tree (DT), SVM, and KNN, to automate PCOS detection. It can be noticed that the proposed LR beat other algorithms in terms of critically classifying datasets.

Prapty and Shitu [25] offered a method that includes a few ML algorithms, such as SVM, KNN, RF, and NB classifiers. NB and RF gave similarly and the best accuracy. Mehrotra et al. [26] suggested a modern approach to PCOS screening. Two classifiers, namely Bayesian and LR classifiers, are used to image processing features extraction to determine a cyst or not. For feature extraction watershed algorithm is used, and for measuring parameters, OpenCV is used. The model can

give consistent results. Denny et al. [27] proposed a system for the early discovery and forecast of PCOS from an optimal and negligible but promising scientific and metabolic parameter. Classification of PCOS with the feature set transformed with Principal Component Analysis (PCA) is conducted using various ML techniques such as the NB classifier method, LR, KNN, CART, RF, and SVM. Bharati et al. [28] showed that several classifiers such as gradient boosting, LR, RF, and hybrid Random Forest and Logistic Regression (RFLR) are applied to the dataset to detect PCOS. In the beginning, a single attribute selection method is used to identify the essential traits for predicting PCOS. Following that, the dataset is split using holdout and cross-validation procedures. At last, the dataset is tested with a variety of classifiers. RFLR has the highest testing accuracy of 91.01% and the highest recall rate of 90% when 40-fold cross-validation is performed to the ten most essential characteristics. Deshpande and Wakankar [29] proposed an automated detection of PCOS by calculating the number of follicles in an ovarian ultrasound picture and then consolidating clinical, biochemical, and imaging parameters to classify patients in two bunches, i.e., normal and PCOS-influenced. The model consists of preprocessing, feature selection, segmentation, and classification. For classification, the SVM approach is applied. Alagarsamy et al. [30], proposed an approach to detect PCOS that uses a heat map preprocessing step. After preprocessing, the data is used for training and testing. The authors used SVM, NB, KNN, and ensemble methods for training and classification. Ensemble methods include SVM and KNN. The Ensemble method outperformed with an f1-score of 98.06452%.

Nsugbe [31] offered a method of detecting PCOS that uses SMOTE to make the imbalanced dataset into a balanced one. Ten different ML classification algorithms like DT, Linear Decision Analysis (LDA), LR, KNN, Liner SVM (LSVM), Quadratic SVM (QSVM), Cubic SVM (CSVM), Fine Gaussian SVM (FGSVM), Medium Gaussian (MGSVM), Coarse Gaussian SVM (CGSVM) have been used for analyzing the performance. SVM gave the best result among all the classifiers. Hari et al. [32] suggested a multi-disease detection model by using six ML models namely LR, RF, KNN, SVM, NB, and voting classifier (Boosting). They detected thyroid, PCOS, and liver disease by using this model. The suggested model has four parts data acquisition, data cleaning, attribute subset selection, and feature scaling. Data is acquired from the Kaggle dataset. After acquisition, data is cleaned to fill in the missing values, smooth the noisy data, etc. Then the dataset subset is selected. For dataset subset selection Recursive Feature Elimination (REF) and Analysis of Variance (ANOVA) are used. To standardize the independent variables of a dataset feature scaling is used. Khanna et al. [33] evaluated a method where 12 distinct machine learning models—LR, DT, RF, SVM, NB, KNN, AdaBoost, XGBoost, and Extratrees—were evaluated and examined. To develop a model to test for PCOS, they used an ensemble learning strategy. Three stacks were created. The LR, SVM (with linear, polynomial, gaussian, and sigmoidal kernels), NB, and KNN models were among the seven classifiers that STACK-1 combined. STACK-2 stood in for an ensemble of tree-based classifiers that included DT, RF, AdaBoost, XGBoost, and Extratrees. STACK-3 was constructed by stacking STACK-1 and STACK-2 multiple levels high. With accuracy, precision, recall, and F1-score of 98%, 97%, 98%, and 98%, the final multi-stack of ML models performed best.

Vasavi et al. [34] developed a model that detects PCOS using classification algorithms. They also developed an app for PCOS victims and finally established a hardware requirement. For detecting PCOS, they used DT, KNN, LR, NB, SVM, and RF. Before using the algorithms, the data was preprocessed, and the feature was selected. Aggarwal and Pandey [35] aggregated obesity, diabetes, high blood pressure, and heart disease for identifying PCOS using supervised, and unsupervised ML techniques. For supervised techniques, they used RF, DT, Gradient Boosting (GB), KNN, LR, SVM, and Hybrid RF and Logistic Regression (HRFLR). For supervised techniques, they used the k-means clustering algorithm.

Hdaib et al. [36], using MATLAB, created a high-performing diagnostic model. They used seven classifiers to implement several machine algorithms. Results showed that the KNN classifier performed best in terms of sensitivity, whereas the linear discriminant classifier performed best in terms of accuracy. Additionally, a comparison with four other research publications that used the same PCOS dataset and used the same implementation platforms, evaluation methods, classifiers, and classes, as well as a comparison of each classifier's accuracy and precision, was made. Mehr and Polat [37] used ensemble, traditional and applied classifiers to diagnose PCOS using the Kaggle PCOS dataset. The dataset with all features and reduced subsets of features was created using the filter, embedding, and wrapper feature selection methods. It was used to investigate several classifiers, including Ensemble RF, Extra Tree, Adaptive Boosting (AdaBoost), and Multilayer Perception (MLP). Kyrou et al. [38] showed that women with PCOS regarding the potential risks from COVID-19 and how this may affect their management is also essential. It raises the likelihood of COVID-19-related complications. Maintaining a high quality of care for complicated patients, as with many other women with PCOS, and providing relevant practical suggestions should be prioritized.

Extreme Boosting and RF were coupled by Bhat and Gupta [39]. For early detection of PCOS, they proposed a new technique incorporating the XGBRF and CatBoost models. To fully support this successful categorize performance, data were re-sampled using SMOTE to eliminate outliers and data imbalance. To compare the findings, different classifiers such as Gradient Boosting, RF, logistic regression, HRFLR, SVM, DT, and MLP were used as baseline techniques. CatBoost and XGBRF outscored all other models, with accuracy scores of 95% and 89% on the top ten parameters, respectively. Inan et al. [40] planned a system to classify PCOS by using Extreme Gradient Boosting (XGBoost). It is a DT-based ensemble ML algorithm. At first, two feature selection algorithms are used, namely the Chi-Square Test and the Analysis of variance (ANOVA) test. Then SMOTE and Edited Nearest Neighbors (ENN) are induced to sample data, and at last, XGBoost is used to detect. Alshakrani et al. [41] suggested a mechanism for detecting PCOS that uses HRFLR, combined Extreme Boosting with RF, Linear SVM (LSVM), Light Gradient Boosting Model (LGB), and CatBoost model. Accuracy, precision, recall, f1-score, ROC curve plot, Area Under the Curve Score, and K-fold Cross Validation (CV) evaluation metrics were used to test each model. Finally, results from discussions and comparisons show that CatBoost performs better than competing models, with a 92% accuracy rating.

Gupta et al. [42] presented discriminant investigation in different measurements with Linear and Quadratic shapes for binary classification along with measurements. Training accuracy came to 97.37%, and testing exactness of 95.92% using Quadratic Discriminant Analysis. More improvements can be made in the analysis area. In order to identify PCOS using patient symptom data, a modified ensemble machine learning (ML) classification approach that uses five conventional ML models as base learners and one bagging or boosting ensemble ML model as the meta-learner of the stacked model is proposed. Suha and Islam [43] suggested an approach to choose various sets of features with different numbers and combinations of attributes. Three basic types of feature selection procedures are used. Table 1 summarizes the existing works based on objectives, applied algorithms for fulfilling the objectives, loopholes in that work, and the remarks of the authors of this work. As a summary of all works is described in this table, researchers will be able to get all the information briefly. Moreover, the remarks will help researchers to conduct their research effectively.

From the above discussion, some major points can be found. The existing works did not provide a detailed concept of traditional PCOS detection methods. They discussed partially the techniques. The researchers did not give a proper discussion about existing datasets other than their used ones. They did not find how the datasets are performing in various ML models. Moreover, they did not compare the performances of other literatures with theirs. The challenges and future direction for detecting PCOS using ML are not precisely elaborated.

III. EXISTING PCOS DETECTION TECHNIQUES

Various types of PCOS detection techniques, their parameters, and their structures are described in this section. PCOS detection has two major groups of parent detection techniques. One is the traditional detection process, and the other is ML-based detection techniques. In the traditional detection techniques, the normal and PCOS hormonal range are very important. In this section, a detailed description of this hormonal range is given inside a table for easy understanding. The structures of ML algorithms are given for clear understanding of the researchers. A figure is included that clearly points out all the detection techniques of PCOS. Also, a detailed figure is given regarding the traditional detection process of PCOS, how doctors utilize traditional methods for PCOS detection.

A. TRADITIONAL METHODS

1) HORMONAL TEST AND SYMPTOM AGGREGATION

For PCOS detection, some hormone levels are considered that are:

i. Luteinizing Hormone (LH) and Follicle Stimulating Hormone (FSH): These two are important hormones for the female body. LH concentrations rise to around 25-40 mlU/ml before 24 hours of ovulation. Often women experiencing PCOS have LH levels of around 18mlU/ml and FSH levels of approximately six mlU/ml. This was once thought to be an essential factor in PCOS diagnosis.
ii. Testosterone: Overall, testosterone consists of the sum of all testosterones in the human body, including free testosterone. This ranges from 6.0 to 86 ng/dl. The quantity of testosterone in your body that is unbound and active is referred to as free testosterone, and the quantity is normally between 0.7 and 3.6 pg/ml. Both total and free testosterone levels are frequently elevated in women with PCOS.
iii. DHEA-S: These levels in most PCOS women are greater than 200 ug/dl.
iv. Prolactin: Women with PCOS have an increased rate of prolactin, often between 25 and 40 ng/ml.
v. Estrogen: Many women having PCOS have normal estrogen levels about 25-75 pg/ml).
vi. Thyroid Stimulating Hormone (TSH): The level of TSH in PCOS women is normally (0.4-3.8 uIU/ml) [44].

Table 2 describes the range of normal and PCOS affected hormones. This table is useful for researchers to understand traditional PCOS detection methods.

Along with the hormone levels, various symptoms are considered for PCOS detection. Some symptoms of PCOS detection are obesity, irregular menstrual, excessive hair fall, an increase in male hormones, etc. Physicians check the hormone level and consider the symptoms for detecting PCOS manually. The hormonal test is the most expensive and is a lengthy process to detect PCOS. A set of questionnaires is used to detect the symptoms of PCOS, traditionally performed. Questionnaires could be used to detect clinically obvious PCOS within relatives of PCOS patients. Though interviewing with a written questionnaire can find a high majority of afflicted mothers, approximately 50% of sisters with PCOS remain unreported [45].

TABLE 1. The following table summarizes the objectives, used algorithms, loopholes, and significant remarks of existing PCOS detection works.

Reference	Objective	Used Algorithm	Loopholes	Remarks
Sumathi *et al.* [10], 2021	To automate PCOS detection	CNN	Does not have a proper description of the dataset.	Watershed segmentation algorithm is used which gives high accuracy. CNN can be used for segmentation as it is very fast.
Cahyono *et al.* [11], 2017	To detect PCOS and give robustness of the system to handle real-world data	CNN	Testing performance is relatively low	In this research, the combination of dropout rate 0.1 and learning rate 10^-6 shows the best performance. But other combination can be tried to achieve better performance.
Dewi *et al.* [12], 20	Automatic detection of PCOS	CNN	The accuracy rate is not very high	This model uses single layer. More hidden layers can be added. More layers can increase the efficiency of the model, but it may decrease the accuracy.
Hossai *et al.* [13], 2022	To detect polycystic ovary from ovarian ultrasound images	CNN	Small dataset	The model is composed of 45 layers. It runs for 30 epochs with a learning rate of 0.00001. So, the total time for running the model needs to be considered.
Bhosale *et al.* [14], 202	The work uses DCNN-based image processing feature extraction to classify PCOS in the dataset.	DCNN	The proposed model does not mention any performance percentage rate.	For feature selection, filter based univariate feature selection is used. But embedded feature selection method can be used as it has both the qualities of filter method and wrapper method
A. Alamoudi *et al.* [15], 2023	To diagnosis PCOS from a data set that includes the ultrasound image of the ovary with clinical data	VGG16 VGG19 Inception v3 MobileNet DenseNet 121 DenseNet 201	The number of samples in dataset. The dataset is used from a single center. Lack of available computational resources	The images where the ovaries were clearly visible, were considered. But noisy images should also be considered.
Thara *et al.* [18], 2021	The goal of this study is to use SNSSDA -ANN-driven feature extraction to determine if an ovary is normal or has PCOS using ultrasound images.	ANN and adaptive k-means clustering algorithms	Not mentioned	This approach uses adaptive k-means clustering algorithm for feature selection. Fuzzy c means can also be used. Adaptive k-means is relatively simple but fuzzy c means gives better results for overlapped data sets.
Maheshwari *et al.* [19], 2020	Use furious fies for feature identification and presents an automated and effectual approach for PCOS diagnosis with ovarian images.	NB and ANN	Data is collected from finite samples.	-
Hassan *et al.* [20], 2020	Detecting PCOS based on clinical signs using prominent ML algorithms on random data sets.	SVM, CART, NB Classification, RF, and LR	Small dataset	In this work, data preprocessing is not described in detail. Rather cross validation can be used to make the model more efficient.
Nandipati *et al.* [21], 2020	To determine which categorization model and attributes are important in illness prediction	RF, KNN, SVM, NN and NB	Not mentioned	5 feature selection methods were implemented and at last correlation value was used for selecting the final features. This approach selects features in an unbiased way.
Thakre *et al.* [22], 2019	For early-stage detection and prediction	RF, SVM, LR, Gaussian NB, KNN	Dataset is small	From the ROC curve, it is visible that the AUC is greater than the red dotted line. That means the model performs well. But it can be made better.
Dutta *et al.* [24], 2021	To learn about the best ML classification algorithm	LR, RF, DT, SVM and KNN	Small dataset	SMOTE is used for balancing the dataset. The improved version of SMOTE, ADASYN (Adaptive Synthetic) can be used as it makes the samples more realistic.
Prapty *et al.* [25], 2020	To improve PCOS detection performance	SVM, KNN, RF, and NB	Put emphasis on a smaller number of attributes.	The work focuses on 7,8 attributes among 31 attributes. More attributes can be incorporated.

TABLE 1. ***(Continued.)*** **The following table summarizes the objectives, used algorithms, loopholes, and significant remarks of existing PCOS detection works.**

Mehrotra *et al.* [26], 2011	To statistically evaluate metabolically and clinical features based on probability density function and box plot.	NB and LR	Accuracy is not up to the mark.	-
Bharati *et al.* [28], 2020	To screen individuals at an early stage in the disease to avoid any significant complications.	Gradient boosting, RF, HRFLR	Less number of features are used.	40 k-fold cross validation increases the model accuracy, but the computational time should also be considered.
Deshpande *et al.* [29], 2014	Automated detection of PCOS. Feature extracted using Multiscale morphological approach.	SVM	The dataset used is very small.	For segmentation canny edge detection technique is used. It produces smooth edges, but it is quite time consuming. Sobel operator can be used as it is time efficient, but the edges are rough.
M. Alagarsamy *et. al* [30],2023	To predict PCOS fertility and infertility	SVM, KNN, NB and ensemble (SVM+KNN)	Less number of classifiers have been used.	The error rate for SVM and NB is relatively low. These two models can be ensembled together to get better performance.
Ejay Nsugbe et al. [31],2023	To detect PCOS	DT, LDA, LR, KNN, LSVM, QSVM, CSVM, FGSVM, MGSVM, CGSVM	Sample size and its nature as it is an unbalanced dataset.	Probabilistic SVM classification model was invoked in this work along with binary classification model that brings a new approach.
D. N. S. Hari *et al.* [32],2023	To identify thyroid, PCOS, and liner disease	LR, RF, KNN, SVM, NB, and voting classifier (Boosting)	Detects only if the data is in numerical format. It cannot be used as a replacement for the present system.	Introducing cross validation might improve the model efficiency.
V. V. Khanna *et al.* [33], 2023	To accurately detect PCOS, to assist medical professionals in decision-making.	LR, DT, NB, RF, SVM, KNN, AdaBoost, XGBoost, ExtraTress, STACK1, STACK2, and STACK3	The dataset needed is not of high quality.	The AUC curve almost touches the peak point. That means the model performs well.
R. R. Vasavi *et al.* [34], 2023	To build an application to monitor women's menstruation conditions	DT, KNN, LR, NB, SVM, and RF	The app and hardware infrastructure need to be developed more.	The proposed work not only detected PCOS, but also developed an app to maintain the lifestyle. But it would have been better if the feature selection method was described vividly.
S. Aggarwal *et al.* [35], 2023	To find disorders that are similar to PCOS that can help with early detection	RF, DT, GB, KNN, LR, SVM, HRFLR and K-means	Not mentioned	Authors combine different datasets to create a new one.
Hdaib *et al.* [36], 2022	To form a new and easy method to diagnose PCOS	KNN, NN, NB, SVM, DT, LR, LDA	Not mentioned	Detailed description of the feature selection method and preprocessing step is required for better analysis.
Mehr *et al.* [37], 2021	To detect PCOS early	Ensemble RF, Extra Tree, Adaptive Boosting (AdaBoost), and MLP	Not mentioned	-
Bhat *et al.* [39], 2021	To improve the detection performance	Gradient Boosting, RFLR, HRFLR, SVM, DT, and MLP	The model fails to predict type-2 error, and a standard number of features are needed to get the desired result.	For high cross validation, the standard deviation is also high. That means the data is widely spread. A low cross validation can be considered. But in that case the accuracy will be a little low.
S. Alshakran *et al.* [41], 2022	Early detection of PCOS	LSVM, LGBM, XGBRF, HRFLR, CatBoost	Limited data	-

TABLE 2. Normal range and the change of hormones in PCOS, along with the detected method.

Feature	Parameters		Method
	Normal range	Range in PCOS	
FSH	4.7-21.5 IU/L	Increased level.	Blood test
LH	5-25 IU/L	Increased level.	Blood test
Prolactin	Less than 25 ng/mL	Normal	Blood test
DHEA	35-430 μg/dL	Greater than 200	Blood test
Testosterone	15-70ng/dL	Increased level. Mostly less than 150ng.dL	Blood test
Estrogen	30-40pg/mL	Level falls	Blood test

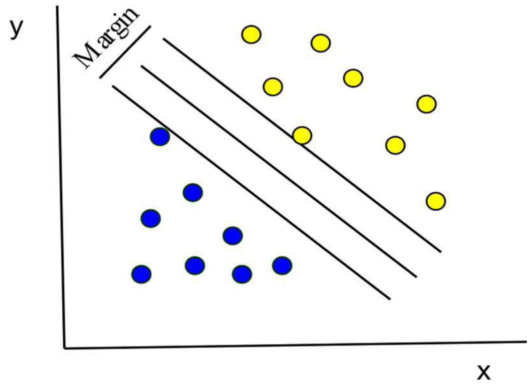

FIGURE 6. Visual representation of the working approach of SVM. Margin separates the two different classes generated by SVM.

2) MANUAL ULTRASOUND IMAGE

Doctors manually detect PCOS based on the number of cysts from the ultrasound image. The presence of more than 12 follicles measuring 29 mm in each ovary may indicate PCOS [46]. Doctors manually count the cysts and detect PCOS. It is time-consuming. Moreover, errors can occur, for example, a mistake during counting, not considering any cyst by mistake, considering any lump as a cyst, etc. Except for the abdomen ultrasound, another ultrasound is available. It is transvaginal ultrasound. In this process, a lubricated probe is inserted inside the vagina of the female, and the doctor sees the inside situation of the organs through a monitor. They analyze the situation of the uterus and cervix and manually count follicles on the ovaries. They also measure the volume of ovaries to detect PCOS. But this process is not error-free. Doctors can make mistakes while counting follicles in the ovaries in measuring the volume of ovaries. Most importantly, this is not a painless method, and many people do not prefer this to social and cultural barriers.

B. MACHINE LEARNING (ML)

ML is a growing field increasing area of processing techniques that targets to mimic human intelligence by learning from their surroundings [47]. ML can be divided into two basic parts classification and clustering.

1) CLASSIFICATION

A supervised data mining technique is classification. This technique is used to classify a fresh observation based on its structure into existing categories or classes. A classification model is used to identify these categories, which lets us estimate the group IDs or class labels of unseen data classes with unknown labels [48]. The classification algorithms are given below:

i. Support Vector Machine (SVM): It is a classification and regression prediction algorithm that employs ML theory to enhance predictive accuracy while automatically avoiding over-fitting to the data. SVM is a valuable data classification technique. SVM creates a hyper line that separates data into different classes. It can solve linear as well as non-linear problems. It is also applicable for real life problems. SVM's main advantage is that training is reasonably simple. Unlike neural networks, there is no local ideal. One weakness is the requirement for a good kernel function [49]. Fig. 6 represents the activity of the SVM algorithm. This figure shows how SVM algorithm classifies data by putting a margin line. About 15 works used SVM for PCOS classification. Sometimes this algorithm is used for ensemble with other algorithms. SVM performed very well compared to other literatures. From Satish et al. [21], it can be seen that SMOTE based SVM performed well in case of average precision, recall and f1-score. Nsugbe [31] used higher variations of SVM like LSVM, QSVM, CSVM, FGSVM, MGSVM, CGSVM.

ii. Naïve Bayes (NB): It is a supervised learning algorithm based on Bayes theorem. This classifier substantially improves learning by presuming that attributes are irrespective of class. In this algorithm, each pair of features being categorized is independent of each other. In practice, NB usually beats more complicated classifiers, even though independence is a weak condition in general [50].
Despite having demerits, this classification algorithm is used by approximately 9 research works to detect PCOS. NB is used with NN and with other classification algorithms. Prapty and Shitu [25] used it and achieved 93% accuracy which was the second highest accuracy after RF.

iii. K-Nearest Neighbors (KNN): This approach is a basic but effective classification method. It is a supervised learning method that uses proximity to classify data. It works on the basis of finding distance between a fixed point and other existing points.
The KNN classification method is a non-parametric classification technique that is basic but beneficial in

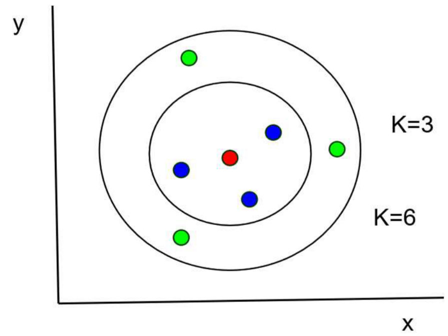

FIGURE 7. **Representation of KNN. The red point is the origin. The blue points are classified when k=3, and when k=6, both the green and blue points are classified.**

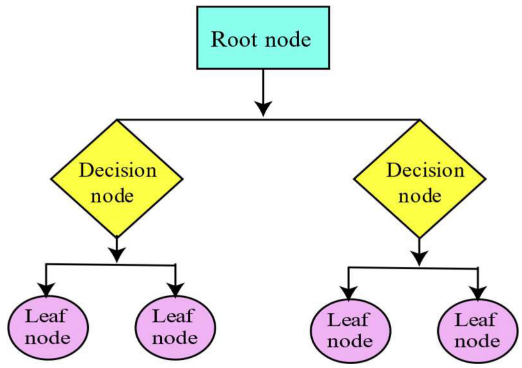

FIGURE 8. **Working principle of a decision tree. The process starts at the root node. Decision nodes make leaf nodes and make various classes.**

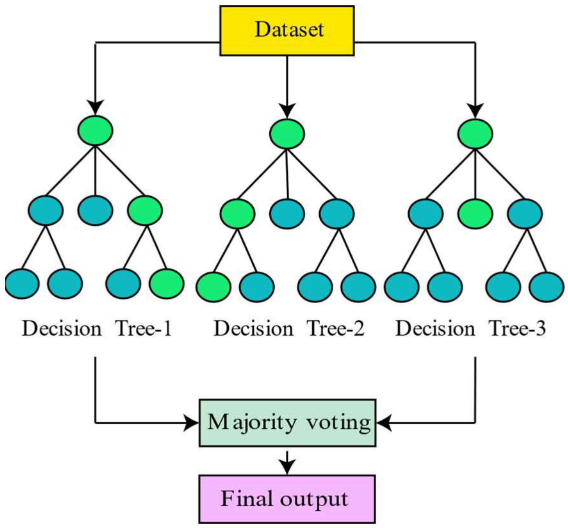

FIGURE 9. **Visual presentation of random forest. Various decision trees are generated, and through majority voting, the final output is found.**

several circumstances. For a data record "t" to be categorized, its "k" nearest neighbors must be found. Neighbors are obtained, resulting in a neighborhood of "t". The majority vote among the data records in the neighborhood is typically used to determine t's classification with or without considering distance-based weighting. However, to use KNN, an acceptable value for k must be selected. In certain ways, the KNN approach is influenced by k [51]. Fig 7 shows the working algorithm of KNN. This gives a vivid understanding of how different classification is comprised in KNN. In this work, various literatures are identified that uses KNN for PCOS identification. Satish et al. [21] used the value of k=10 to optimize KNN. Denny et al. [27] also used KNN with other classifiers where KNN performed relatively well. KNN was utilized by almost 11 research works.

iv. Decision Tree (DT): This is a straightforward yet powerful categorization method. One of its significant benefits is that they give understandable classification criteria to humans. One downside of the DT is that to decide where to separate a node, all quantitative attributes must be sorted. This is time-consuming and memory-intensive, especially when DTs are trained on massive volumes of data [52]. Fig. 8 describes the working mechanism of the DT. This figure shows how decision node makes classification. Though DT has some demerits, still many researchers use it for its powerful capabilities. A total in 7 PCOS detection works DT has been used. Aggarwal and Pandey [35] used DT with other classification algorithms. Most of the recent works optimized DT in their work.

v. Random Forest (RB): This technique was established by L. Breiman in 2001, and it has been a massive success as an overall classification and regression tool. The method, which averages the predictions of multiple randomized DT's, has shown excellent performance in circumstances when the number of variables is significantly more than the set of observations. It is flexible to a range of ad hoc learning exercises and provides measurements of changing relevance, making it suitable for large-scale challenges [53]. Fig. 9 represents the working module of RF. From this figure researchers can visualize how RF works. Almost in 12 PCOS detection works, RF is used a classification tool. Bharati et al. [28] achieved a high accuracy of detection using RF. They also used RF in a hybrid model with LR. This model performed the best result in their work.

vi. Logistic regression (LR): It is a technique for estimating the likelihood of a binary outcome. The often-used logistic regression models produce a binary with two possible values, such as true/false, yes/no, and so on [54]. LR has a high utilization rate in case of PCOS detection. It is also used with RF to build hybrid models. This hybrid model outperforms the other classification in case of PCOS detection.

2) CLUSTERING

Clustering is an ML approach that deals with data point grouping. A set of data points can be clustered based on some shared characteristics. This approach is beneficial for

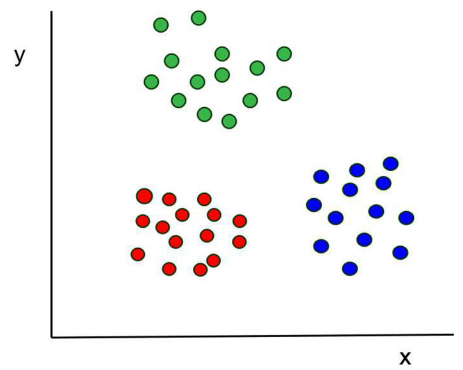

FIGURE 10. Visualization of the clusters generated by k-means algorithm.

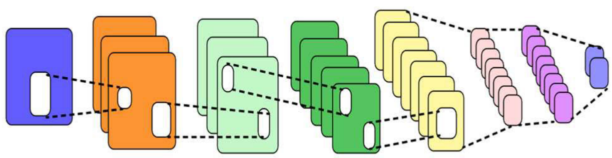

FIGURE 11. The architecture of CNN algorithm having various layers.

detecting anomalies. It is a method of unsupervised data mining [48].

i. K-means: This clustering method separates an unlabeled dataset into clusters. K defines the set of pre-clusters that must be formed during work; like, if K=2, two clusters will be produced; if K=3, three clusters will be produced, and so on [55]. It is the only clustering algorithm that has been used for PCOS detection. It has been only used for two times. S. Aggarwal and Pandey [35] used basic k-means algorithm whereas, Thara [18] used adaptive k- means for PCOS detection. Fig. 10 describes K-means. Readers can easily visualize the different clusters from this figure.

3) ARTIFICIAL NEURAL NETWORK (ANN)

ANN analyzes the human mind's neural network from the perspective of information processing, produces a simple model, and assembles many networks based on various connection A network is an information processing model that consists of several interconnected nodes (or neurons).

Each node represents an activation function, which is a specific output function. The weight of the signal traveling along the connection is indicated by the link between every two nodes and is equivalent to the memory of an ANN. The output of the network is different depending on how it is linked, the weight quantity, and the reward function [56].

i. Convolutional Neural Network (CNN): A basic deep neural network is CNN. It is a linear mathematical procedure between two integers. CNN has many layers, including the layers that are convolutional, nonlinear, pooling, and fully connected layers that are convolutional and totally coupled; pooling and

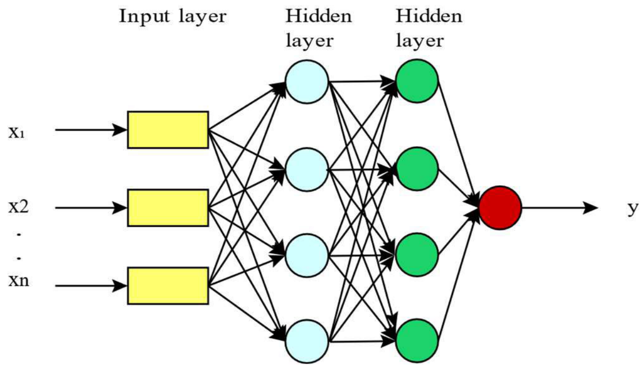

FIGURE 12. Vivid visualization of MLP architecture. The input layer is connected with multiple hidden layers.

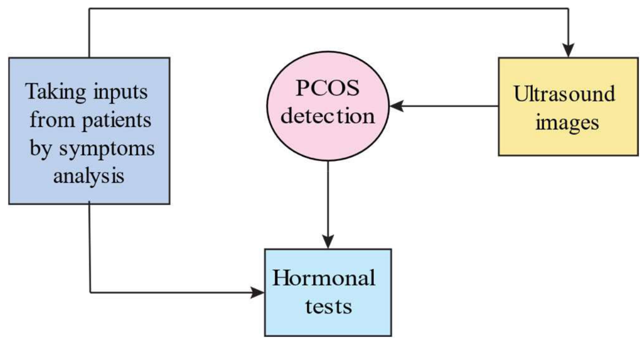

FIGURE 13. The working approach of traditional PCOS detection method.

non-linearity layers do not have parameters. Pooling and non-linearity layers do not have parameters. CNN performs exceptionally well in ML because of its diversity [57]. It is a very efficient algorithm in case of cyst detection from ultrasound images. The CNN trained models can easily predict whether an ovary is normal or PCOS infected. In total 6 works have been discussed in this work that uses CNN for detecting PCOS. These models use different datasets and hidden structures. Most of the models produce good results. Cahyono et al. [11] used CNN architecture to detect PCOS. This model achieved the perfect performance rate 0f 100%. Fig 11 represents the architecture of CNN. The figure clearly represents various layers and their working principle of CNN.

ii. Multilayer Perception: It is a neural network that has a nonlinear input-output alignment. Input and output layers, and one or many hidden layers with many neurons stacked on top of one another, make up an MLP. Despite neurons in a perceptron, which must use a threshold-enforcing activation function like ReLU or sigmoid, neurons in an MLP can use any random activation function [58]. It a very least used ANN technique for PCOS detection. Fig. 12 depicts the architecture of MLP. This figure makes easy for researcherchers to understand the architecture and various layers of MLP.

TABLE 3. Dataset details, description, and performance analysis of the used datasets in the existing works.

Architecture	Reference	Dataset	Description	Implementing Medium	Highest Performance	Comment
Neural network	Sumathi *et al.* [10], 2021	Ultrasoundimages.com and sonoworld.com	Image dataset	Python	85%	From the visualization of the performance, the custom dataset made by Cahayoo et al. performed best with the performance rate of 100%. Hossain et al. used two large datasets and obtained the performance of 98.12% which is quite satisfactory. Though the first one has the best performance rate, the dataset is too small to give accurate result. So, the dataset from Hossain et al. performed best compared to other datasets.
	Cahyono *et al.* [11], 2017	Custom	The author introduced the dataset. 40 non-PCO data and 14 PCO data	Not mentioned	100%	
	Dewi *et al.* [12], 2018	Klinik Bersalin Permata Bunda Syariah, Cirebon	Not mentioned	Not mentioned	80.84%	
	Hossain *et al.* [13], 2022	Kaggle and internet	The database consisted of two discrete datasets: dataset A and dataset B. Dataset A- was utilized for training and validating. It contained a total of 1,346 training photos in addition to 384 validation images. Dataset B included a total of 339 photos, 154 of which were healthy ovarian ultrasound images and 185 were PCOS infected ovarian images.	Not mentioned	98.12%	
	Bhosale *et al.* [14], 2022	Kaggle		Python	Not mentioned	
	A. Alamoudi *et al.* [15], 2023	Custom		Not mentioned	82.46%	
Hybrid approach of ANN and clustering	Thara *et al.* [18], 2021	Not mentioned	Image dataset	Not mentioned.	75%	This approach did not mention anything about the dataset and the performance is no satisfactory.
Hybrid approach of ANN and classification	Maheshwari. *et al.* [19], 2020	Custom	68 ultrasound images	MATLAB environment.	100%	For the hybrid approach of ANN and classification, a custom dataset was used with 68 images. Though the model performed well but a larger dataset is needed.
Hybrid approach of NN and classification	Nandipati *et al.* [21], 2020	Kaggle	The dataset was taken from Kaggle having 541 instances	Python-Scikit Learn package and RapidMiner.	93.12%	The dataset obtained from Kaggle used by Hdaib et al. performed best for NN and classification.
	Purnama *et al.* [23], 2015	Custom	Two datasets are used. Dataset A, 40 images consist of 26 normal images and 14 PCOS-indicated images. Dataset B, 40 images consist of 34 normal images and 6 PCOS images.	Not mentioned.	78.81%.	
	Hdaib *et al.* [36], 2022	Kaggle	PCOS dataset, which was released in 2020 and was obtained from 10 hospitals in Kerala,	Python and MATLAB.	97.6%	

TABLE 3. ***(Continued.)*** **Dataset details, description, and performance analysis of the used datasets in the existing works.**

			India, covers the clinical and physical characteristics of 541 women, 177 of whom had PCOS and 364 of whom were healthy.			
Classification	Hassan *et al.* [20], 2020	Not mentioned	42 independent variables	R-language. R libraries used for the purpose were e1071, CARET, naive Bayes, part, randomForest, klaR, ggplot2.	98%	By using classification algorithm, two datasets got the perfect result. Ejay Nsugbe used a dataset collected from Kaggle. This dataset was imbalanced initially, then was made balanced. Mehr et al. also used a dataset collected from Kaggle. Other datasets also performed well, but these two gave the perfect performance.
	Thakre *et al.* [22], 2019	Kaggle	Owned by Prasoon Kottarathil and has 41 features.	Jupyter Notebook and proposed web app PCOcare is intended to be developed in Python using Streamlit Open-Source Web App Framework.	97%	
	Dutta *et al.* [24], 2021	UCI	UCI PCOS without fertility dataset. The dataset has 541 records.	Python version 3.7.6 with Jupyter Notebook 6.0.3 & Operating system Ubuntu.	98%	
	Prapty *et al.* [25], 2020	Custom	Ten individual hospitals.	Not mentioned.	93.5%	
	Mehrotra *et al.* [26], 2011	Custom	Data were taken from Ghosh Dastidar Institute for Fertility Research (GDIFR). The sample was taken from 200 patients from Kolkata between March 2010 and April 2011.	Not mentioned.	94.23%	
	Denny *et al.* [27], 2019	Custom	The author introduced the dataset having 541 samples.	Python.	98.039%	
	Bharati *et al.* [28], 2020	Kaggle	A dataset with 43 attributes of 541 women is collected from the Kaggle repository	Python programing language. Anaconda distribution package, Scikit-learn library, Jupiter notebook, Spyder, Orange.	91.01%	
	Deshpande *et al.* [29], 2014	Custom	The author Introduced a dataset of 20 patients	Not mentioned.	95%	
	M. Alagarsamy *et al.* [30],2023	Kaggle	The dataset consists of 541 data, among which 364 are infertility data, and 177 are fertility data.	Not mentioned	98.701299%	
	Ejay Nsugbe, [31], 2023	Kaggle	An imbalanced dataset consisting of 364 patients without PCOS and 177 with PCOS. After making balanced, 728 samples were used.	Not mentioned	100%	
	D. N. S. Hari *et al.* [32],2023	Kaggle	Indian dataset is collected from Kaggle. The thyroid disease dataset has 29 attributes, PCOS has 44 features, and liver disease has 11 features	Python	90.1%	
	V. V. Khanna *et al.* [33], 2023	Kaggle	A publicly available dataset that Kottarathil	Python	98%	

TABLE 3. ***(Continued.)*** **Dataset details, description, and performance analysis of the used datasets in the existing works.**

		created on Kaggle contains information on 541 fertile women across 43 attributes. The data samples in this multi-centric dataset come from 10 hospitals in Kerala, India.		
R. R. Vasavi *et al.* [34], 2023	Kaggle	The dataset contains only the main features	Python for data preprocessing, Java for software development, Android Studio and XML for app development	98%
S. Aggarwal *et al.* [35], 2023	Kaggle	This study makes use of two different datasets: one for diabetes (Pima Indians Diabetes Database, n.d.) and one for heart disease (Heart Disease Dataset, 2019). From Kaggle, both datasets were taken.	Jupyter notebook has been used with Python and other programming languages.	98.9%
Mehr *et al.* [37], 2021	Kaggle	Gathered from 10 different Indian hospitals are collected from Kaggle Dataset Repository. The dataset includes 43 features.	Python, Scikit-learn library, and Jupiter Notebook were used.	100%
Bhat *et al.* [39], 2021	Custom	A public dataset published in Kaggle with 45 attributes	Anaconda distribution packages, Jupiter notebook, Scikit-learn library, Matplotlib, are used in Python development.	95%
S. Alshakran *et al.* [41], 2022	Kaggle	The PCOS dataset that Kottarathil published was compiled from ten different hospitals around Kerala, India. The dataset included 44 features based on the physical and clinical characteristics of 541 women, including 177 PCOS-positive women.	Python	95%

Mehr and Polat [37] and Bhat and Gupta [39] have implemented MLP for PCOS detection.

Fig. 13 shows the architecture of traditional PCOS detection approaches. From this figure researchers will easily understand the method of traditional detection. They will be able to compare this process with the other processes. The process starts with taking symptoms from patients. Their subsequent methods are briefly described. It can be seen at first the symptoms are analyzed. Then, hormone tests or ultrasound tests are accomplished based on the symptoms. Based on the result, PCOS is detected. Fig. 14 provides the hierarchical representation of PCOS detection techniques based on traditional and ML methods. The following figure gives an overall overview of all PCOS detection methods. It has summarized the whole PCOS detection process and makes it easier for researchers to do further research.

IV. DATASET

This section gives a detailed overview of the datasets utilized in PCOS detection techniques discussed in section II. It focuses on the dataset's properties, components, and volumes. One of the major aspects of running detection algorithms is the dataset that is used for training and testing.

Although PCOS is a crucial disease for women worldwide, the dataset available for PCOS is minimal. For analyzing the datasets, a table is included in this section. This table summarizes the datasets and compares them on the basis of performance. Table 3 summarizes the dataset used in

TABLE 4. Performance analysis of existing literature based on various performance matrices.

Author	Accuracy	Recall or Sensitivity	Precision	F1-Score	Specificity
Sumathi *et al.* [10], 2021	85%				
Cahyono *et al.* [11], 2017				100%	
Hossain *et al.* [13], 2022	InceptionV3 96.56% PCONet 98.12%				
Maheshwari *et al.* [19], 2020	98.63%	55%	100%	68.76%	100%
Hassan *et al.* [20], 2020	LR 92% SVM 94% CART 90% RF 96% NB 81%	LR 91% SVM 95% CART 94% RF 95% NB 76%	LR 98% SVM 95% CART 92% RF 96% NB 94%	LR 94% SVM 95% CART 84% RF 96% NB 93%	
Dutta *et al.* [24], 2021	97.11%	98%	98%	98%	
Prapty *et al.* [25], 2020	KNN 75% SVM 90% NB 93% RF 93.5%	KNN 66% SVM 82% NB 80% RF 84%	KNN 65% SVM 82% NB 80% RF 84%	KNN 64% SVM 81% NB 80% RF 84%	
Mehrotra *et al.* [26], 2011	NB 93.93% LR 91.04%	NB 82.36 % LR 92.85 %			NB 94% LR 94.23%
Denny *et al.* [27], 2019	LR 85.36% KNN 86.58% CART 82.92% RF 89.02% NB 84.14% SVM 82.92%	LR 64.51% KNN 80.64% CART 83.87% RF 74.19% NB 74.19% SVM 54.83%	LR 98.039% KNN 90.196% CART 82.352% RF 95.833% NB 84.14% SVM 82.92%	LR 95.2384% KNN 83.333% CART 74.285% RF 95.833% NB 82.142% SVM 100%	LR 85.36% KNN 86.58% CART 82.92% RF 89.02% NB 84.14% SVM 82.92%
M. Alagarsamy *et al.* [30],2023	SVM 95.37037% KNN 90.74074% NB 92.59259% Ensemble (SVM+KNN) 97.22222%		SVM 96.103896% KNN 92.207792% NB 93.506494% Ensemble (SVM+KNN) 98.701299%	SVM 96.73203% KNN 93.42105% NB 94.73684% Ensemble (SVM+KNN) 98.06452%	
Hdaib *et al.* [36], 2022	KNN 90.74% Neural Network 85.19% NB 80.56% SVM 90.74% DT 84.26% LR 90.7% LDA 92.60%	KNN 97.22 % NN 90.27% NB 79.16% SVM 91.67% DT 90.27% LR 92.00% LDA 92.20%	KNN 89.74% NN 87.84% NB 90.48% SVM 94.29% DT 86.67% LR 94.54% LDA 97.6%		KNN 77.78% NN 75.00% NB 83.33% SVM 88.89% DT 72.22% LR 87.87% LDA 93.55%
Mehr *et al.* [37], 2021	Ensemble RF 98.89%	100%			
S. Alshakran *et al.* [41], 2022	LSVM 90% LGBM 90% XGBRF 87% HRFLR 91% CatBoost92%	LSVM 81% LGBM 84% XGBRF 74% HRFLR 84% CatBoost 84%	LSVM 90% LGBM 92% XGBRF 91% HRFLR 92% CatBoost 95%	LSVM 86% LGBM 87% XGBRF 82% HRFLR 88% CatBoost89%	

various works published. It describes the dataset along with the source of the dataset. It will help researchers to choose the best dataset for future research work.

V. PERFORMANCE ANALYSIS

In this section, major performance parameters and the performance of various works are discussed comparatively.

Different authors choose different performance indicators to assess performance. The frequently used parameters areaccuracy, recall, precision, f1-score, specificity, AUC, and process time. This section includes tables and figures to analyze the performance of various works. From the tables researchers will be able to compare quantitative performance as well as qualitative performance at a glance. From the figure they will be able to compare the classification algorithms that use the same parameters. Fig. 15 analyses the accuracy, recall, precision, specificity, and f1-score in various literatures using the same classification algorithms. This will

TABLE 5. Qualitative performance evaluation of our studied articles using the Delphi technique.

	Round 1 Question	Round 2 Question	Round 3 Question	
Reference	Detection Type	Is the model has preprocessing step?	Does the model have any language dependency?	Comments
[10]	Simple cyst or PCOS	Yes	Yes	No comment.
[11]	PCO or non PCO	No	No	Incorporating preprocessing step and eliminating language dependency would be great.
[12]	PCO or non-PCO ovary	Yes	No	Making language independent would be great.
[17]	Patients having PCOS of low risk, moderate risk, and high risk.	No	No	It would be better if the preprocessing step is included, and language dependency is removed.
[18]	Ovary as normal or PCOS	Yes	No	Language independence would be better.
[19]	Affected by PCOS or not	Yes	Yes	No comment.
[20]	Affected by PCOS or not.	Yes	Yes	No comment.
[21]	Affected by PCOS or not	Yes	Yes	No comment.
[25]	Affected by PCOS or not	Yes	No	Language independency is appreciated.
[26]	Affected by PCOS or not	No	No	A model with preprocessing step and language independency would be better.
[27]	Affected by PCOS or not	Yes	Yes	No comment.
[28]	Data-driven diagnosis of PCOS infected or not	Yes	Yes	No comment.
[29]	Affected by PCOS or not.	Yes	No	Removing language dependency would be great.
[34]	Affected by PCOS or not	Yes	Yes	No comment.

help the further researchers to compare which model works better on that specific classification algorithm and which does not. Table 4 describes the accuracy, precision, recall specificity, and f1-score of various algorithms in various papers. Researchers will be able to compare the existing works on the basis of performance parameters. It will help in the future research works. Not only quantitative but also qualitative analysis is important. Quantitative analysis gives a numerical value, but qualitative analysis finds the effectiveness that cannot be described in numerical values. Qualitative analysis helps researchers to understand the depth of the research. Table 5 describes the qualitative performance analysis of the existing literature based on Delphi technique. The analysis is based on three round questions. The first question is about the detection type. It means whether the ovary is affected by PCOS or not. The second question is, if the model has any preprocessing step, and the last one is if the model has any language dependency. Based on these questions and the found analysis, the comment is made.

VI. CHALLENGES AND FUTURE DIRECTIONS

This section discusses some of the primary obstacles and difficulties associated with PCOS detection in previous studies to give a roadmap for researchers to examine where the emphasis should be put. Also, future directions are also discussed. With a figure the related difficulties for this work and their probable solutions are discussed. Fig. 16 demonstrates the challenges and their possible solutions. The clear representation of this figure will help the researchers in finding research gaps and conducting future works.

A. INFERIORITY OF STANDARD DATASET

Although a few effective data sets are available, there are some limitations. For example, the dataset available for PCOS detection is exceedingly small, and the datasets are not diverse. Most of the datasets are custom made. The custom datasets are very small. On the other hand, the number of datasets available on Kaggle are very few.

ML works best when the trained and tested dataset is huge, as the model can learn and extract the features well. A large dataset that is broad in perspective and neutral to a particular geography is required. Moreover, a dataset should include women of various ages so that variety is included. If the dataset is not significant and standard, the tested result will not be accurate.

B. IMBALANCED DATASET

A balanced dataset has even types of observations for all classes. The existing datasets are effective, but they are not balanced. One class has a high number of observations, and

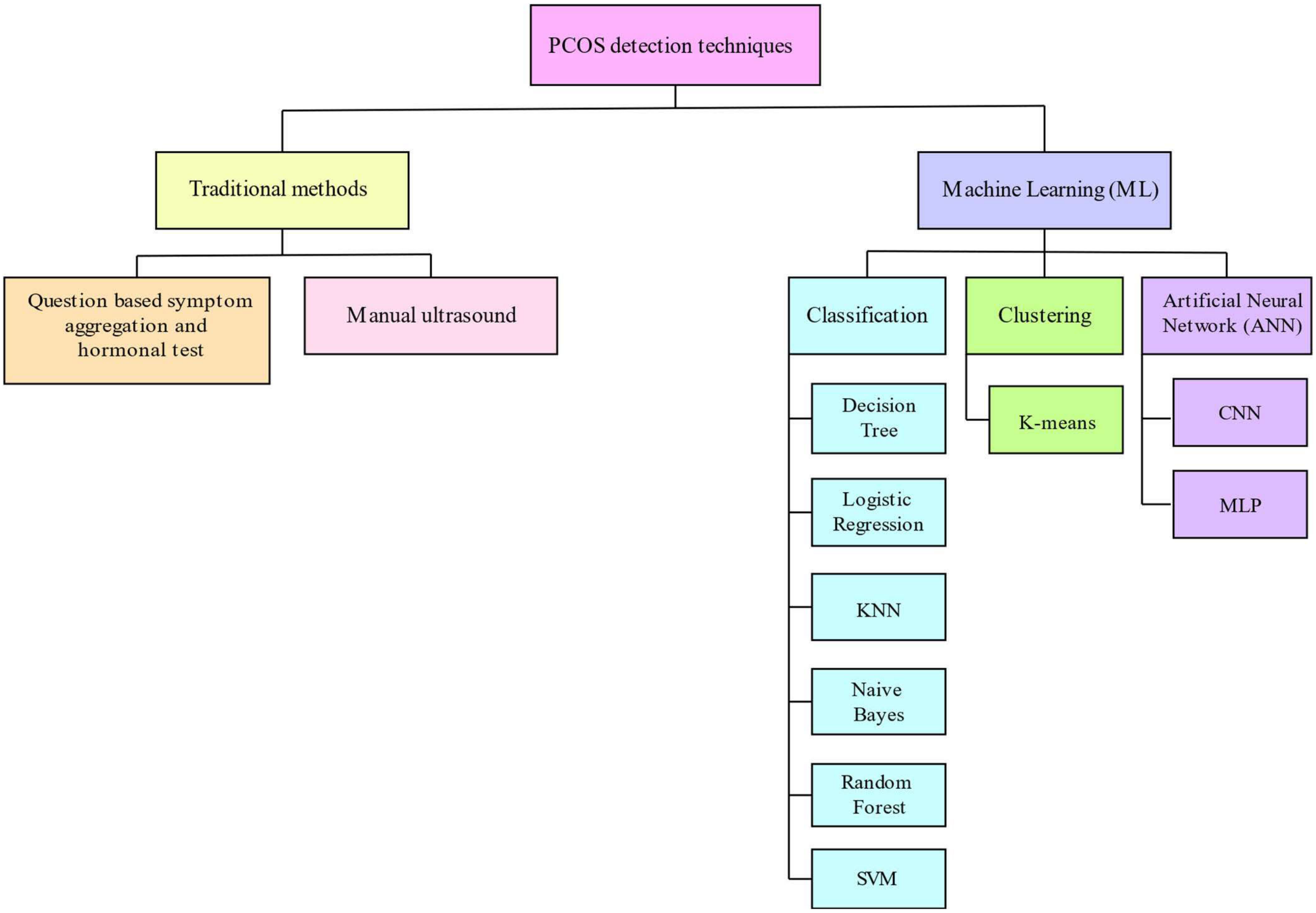

FIGURE 14. **A hierarchical representation of PCOS detection techniques of existing pieces of literature is discussed in section II.**

the other class has a low number of observations. Due to this characteristic, the result is much affected. For this reason, no absolute effect is found. Various preprocessing techniques like PCA and SMOTE can be used to make the dataset balanced.

C. NOISE IN ULTRASOUND IMAGE

To run CNN model, the images must be clear enough, but ultrasound images are prone to speckle noise, salt and pepper noise, and many other noises. These noises must be removed to get the perfect result. As noise is not removed, then the accuracy rate for some CNN models are low.As a result, identification of cysts is not perfect. But in the maximum existing literature, the noise from images is not removed. Therefore, for noise reduction dilation, grayscale, etc., these types of methods need to be used.

D. DETECTION RATE

If researchers want to make PCOS detection automatic and include ML-based techniques in the medical sector, then the detection rate must be 100% so that people can trust machine-based detection instead of manual detection. But in the previous works, detection rate is not perfect. Most of the detection rate is below 98%. This 2% accuracy must be increased by training the model more and more.

E. NOT INCLUDING OBJECT DETECTION

Object detection algorithms have brought a revolutionary change in the field of computer vision. This algorithm has the capability to object localization and object classification at the same time. As a result, object detection algorithms are faster than traditional approaches.

Therefore, using object detection algorithms in detecting PCOS will be very beneficial. The existing literature basically focuses on classifier-based algorithms. But in this sector, YOLO, Fast RCNN, or this type of object detection algorithms can be used for detecting cysts from ultrasound images.

F. LESS USE OF CLUSTERING APPROACHES

For detecting PCOS, only one clustering approach is used, which is the k-means algorithm. There are so many clustering algorithms like DBSCAN, Gaussian mixture model, BIRCH, OPTICS, etc., available. But these algorithms are not

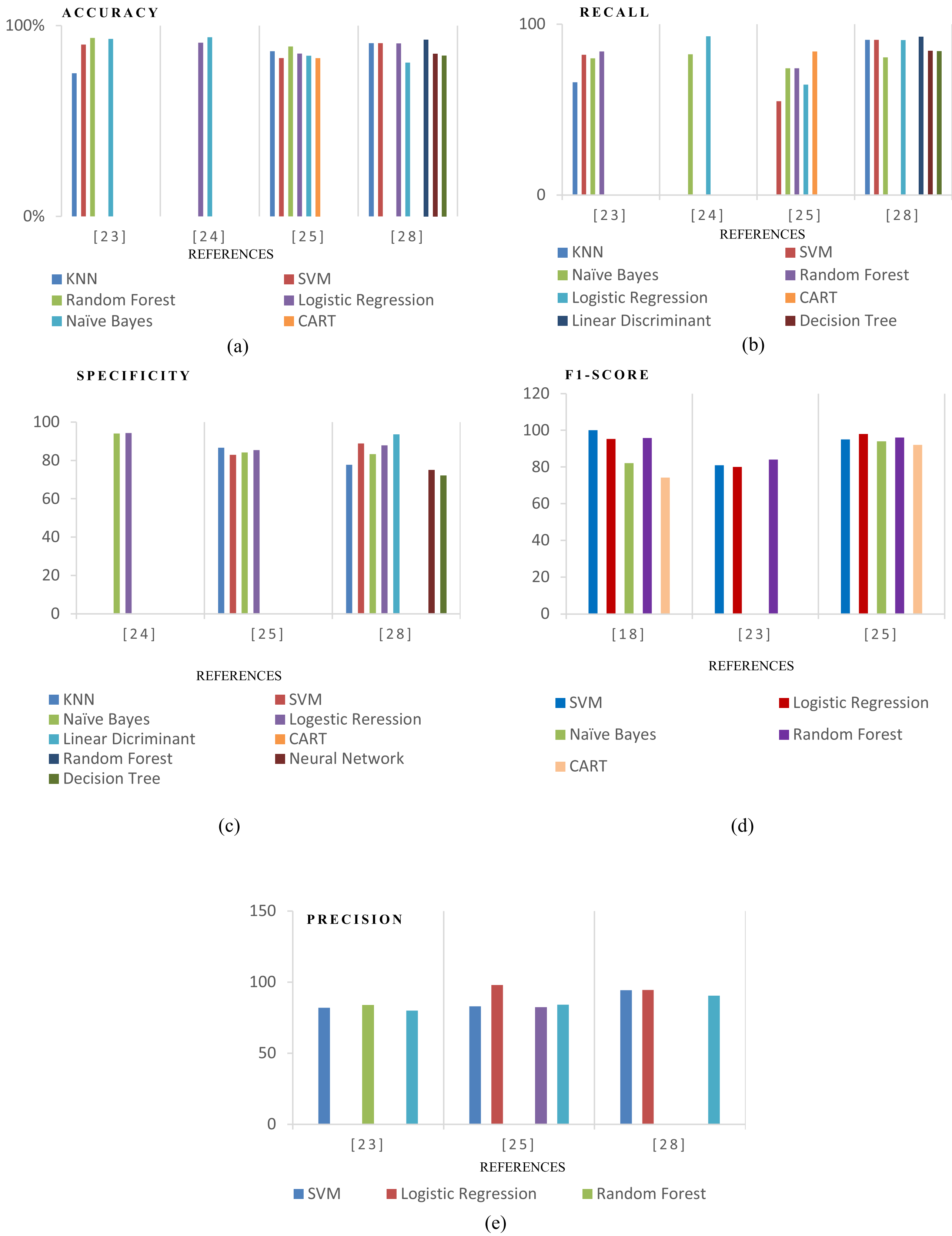

FIGURE 15. Comparison of various performance parameters (a) accuracy, (b) recall, (c) specificity, (d) f1-score, and (e) precision of existing kinds of literature for different classification algorithms.

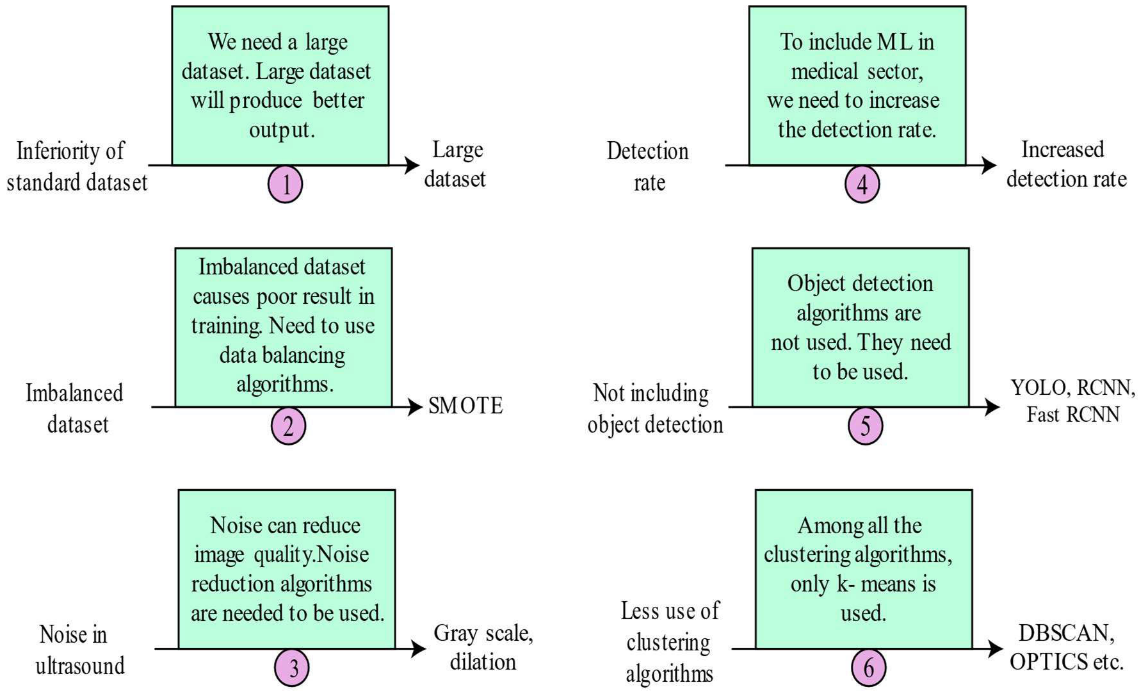

FIGURE 16. Challenges faced in previous literature and their possible solution.

used. These algorithms must be used to find how clustering based models work in detecting PCOS. Therefore, using other clustering approaches rather than K-means can bring more success.

VII. CONCLUSION

This research provided a descriptive and conceptual assessment of all known PCOS detection techniques, focusing on ML. The method of existing algorithms was provided, as well as their aspects, effectiveness, analysis methodology, and outputs. Additionally, the datasets used in such algorithms were shortly described. The flaws of existing algorithms were discussed, as well as potential problems. Though a significant amount of research has been conducted to establish an efficient PCOS detection model, some problems remained unsolved. This paper detected the shortcomings, namely the small number of datasets, imbalance dataset, detection rate, not including more clustering approaches and so on. In future work, we wish to work on a larger dataset having balanced data. Also, much CNN-based optimization needs to be conducted. We would like to use other clustering approaches like DBSCAN, OPTICS rather than K-means. This article will open a new window for the research community to understand the existing ML-based PCOS detection algorithms. By understanding and analyzing the lackings and future scopes in this field, researchers will be able to develop new approaches to solve this problem.

In future work, we would like to study the new PCOS detection algorithms and apply these algorithms to a standard dataset to analyze the performance. We would like to use other clustering algorithms and understand the success rate.

This approach will make it easier to understand the major differences between the new and the old algorithms of ML for PCOS detection.

REFERENCES

[1] B. W. Donesky, "Polycystic ovary syndrome (PCOS)," in *Encyclopedia of Endocrine Diseases*. Amsterdam, The Netherlands: Elsevier, 2004, pp. 1–3.

[2] *Polycystic Ovary Syndrome*. Accessed: May 16, 2022. [Online]. Available: https://www.nhs.uk/conditions/polycystic-ovary-syndrome-pcos/

[3] S. Watson. (Apr. 18, 2021). *Polycystic Ovary Syndrome (PCOS): Symptoms, Causes, and Treatment*. Accessed: May 16, 2022. [Online]. Available: https://www.healthline.com/health/polycystic-ovary-disease

[4] *What's the Treatment for PCOS?* Accessed: May 16, 2022. [Online]. Available: https://www.webmd.com/women/treatment-pcos

[5] V. Gotte and S. Jyothi. (Jul. 5, 2017). *Polycystic Ovary Syndrome Detection Using Various Machine Learning Methods—A Review*. Accessed: Nov. 26, 2022. [Online]. Available: https://www.researchgate.net/publication/339973640_Polycystic_Ovary_Syndrome_Detection_Using_Various_Machine_Learning_Methods-A_Review

[6] J. Vaswani, H. Mulchandani, R. Vaghela, and P. Rajan. *A Systematic Literature Review on Diagnosis of PCOS Using Machine Learning Algorithms*. Accessed: Nov. 26, 2022. [Online]. Available: https://git.org.in/GIT_JET/Papers/Regular%20Edition/8_GIT_JET_Regular_Edition_2022.pdf

[7] S. J. Pulluparambil and S. Bhat, "Medical image processing: Detection and prediction of PCOS—A systematic literature review," *Int. J. Health Sci. Pharmacy*, vol. 5, pp. 80–98, Dec. 2021, doi: 10.47992/ijhsp.2581.6411.0075.

[8] I. O. Rabiu, A. D. Usman, and A. M. S. Tekanyi, "A review on computer assisted follicle detection techniques and polycystic ovarian syndrome (PCOS) diagnostic systems," *Int. J. Comput. Trends Technol.*, vol. 28, no. 1, pp. 41–45, Oct. 2015, doi: 10.14445/22312803/IJCTT-V28P109.

[9] R. Ambad, A. Agrawal, R. Lahoti, P. Muley, and P. Pande, "Role of artificial intelligence in PCOS detection," *J. Datta Meghe Inst. Med. Sci. Univ.*, vol. 17, no. 2, p. 491, 2022. [Online]. Available: http://www.journaldmims.com/article.asp?issn=09743901;year=2022;volume=17;issue=2;spage=491;epage=494;aulast=Agrawal

[10] M. Sumathi, P. Chitra, R. Sakthi Prabha, and K. Srilatha, "Study and detection of PCOS related diseases using CNN," *IOP Conf. Ser., Mater. Sci. Eng.*, vol. 1070, Nov. 2021, Art. no. 012062.

[11] B. Cahyono, M. S. Mubarok, and U. N. Wisesty, "An implementation of convolutional neural network on PCO classification based on ultrasound image," in *Proc. 5th Int. Conf. Inf. Commun. Technol. (ICoIC)*, May 2017, pp. 1–4.

[12] R. M. Dewi, Adiwijaya, U. N. Wisesty, and Jondri, "Classification of polycystic ovary based on ultrasound images using competitive neural network," *J. Phys., Conf.*, vol. 971, Mar. 2018, Art. no. 012005.

[13] A. K. M. S. Hosain, M. H. K. Mehedi, and I. E. Kabir, "PCONet: A convolutional neural network architecture to detect polycystic ovary syndrome (PCOS) from ovarian ultrasound images," 2022, *arXiv:2210.00407*.

[14] S. Bhosale, L. Joshi, and A. Shivsharan. *PCOS (Polycystic Ovarian NDROME) Detection Using Deep Learning*. Accessed: May 16, 2022. [Online]. Available: https://www.irjmets.com/uploadedfiles/paper/issue_1_january_2022/18221/final/fin_irjmets1641558545.pdf

[15] A. Alamoudi, I. U. Khan, N. Aslam, N. Alqahtani, H. S. Alsaif, O. Al Dandan, M. Al Gadeeb, and R. Al Bahrani, "A deep learning fusion approach to diagnosis the polycystic ovary syndrome (PCOS)," *Appl. Comput. Intell. Soft Comput.*, vol. 2023, pp. 1–15, Feb. 2023.

[16] P. Chitra, K. Srilatha, M. Sumathi, F. V. Jayasudha, T. Bernatin, and M. Jagadeesh, "Classification of ultrasound PCOS image using deep learning based hybrid models," in *Proc. 2nd Int. Conf. Electron. Renew. Syst. (ICEARS)*, Mar. 2023, pp. 1389–1394.

[17] V. Krishnaveni, C. Deepa, R. S. Cindhu, and K. G. Santhiya. *An N Based Screener for the Early Diagnose of Polycystic Ovarian NDROME in Adolescent and Young Women*. Accessed: May 2022. [Online]. Available: http://www.joics.net/images/full_pdf/1584184400_M196.pdf

[18] L. Thara and T. M. Divya, "Detection and prediction system for polycystic ovary syndrome using structural normalized square similarity detection approach," *Nveo-Natural Volatiles Essential Oils J.*, vol. 8, no. 6, pp. 2834–2842, 2021.

[19] K. Maheswari, T. Baranidharan, S. Karthik, and T. Sumathi, "Modelling of F3I based feature selection approach for PCOS classification and prediction," *J. Ambient Intell. Humanized Comput.*, vol. 12, no. 1, pp. 1349–1362, Jan. 2021.

[20] M. M. Hassan and T. Mirza, "Comparative analysis of machine learning algorithms in diagnosis of polycystic ovarian syndrome," *Int. J. Comput. Appl.*, vol. 175, no. 17, pp. 42–53, Sep. 2020.

[21] C. R. N. Satish, X. Chew, and K. W. Khaw, "Polycystic ovarian syndrome (PCOS) classification and feature selection by machine learning techniques," *Appl. Math. Comput. Intell. (AMCI)*, vol. 9, pp. 65–74, Jan. 2020.

[22] V. Thakre, "PCOcare: PCOS detection and prediction using machine learning algorithms," *Biosci. Biotechnol. Res. Commun.*, vol. 13, no. 14, pp. 240–244, Dec. 2020.

[23] B. Purnama, U. N. Wisesti, F. Nhita, A. Gayatri, and T. Mutiah, "A classification of polycystic ovary syndrome based on follicle detection of ultrasound images," in *Proc. 3rd Int. Conf. Inf. Commun. Technol. (ICoICT)*, May 2015, pp. 396–401.

[24] P. Dutta, S. Paul, and M. Majumder, "An efficient SMOTE based machine learning classification for prediction & detection of PCOS," Tech. Rep., 2021. [Online]. Available: https://www.researchsquare.com/article/rs-1043852/v1

[25] A. S. Prapty and T. T. Shitu, "An efficient decision tree establishment and performance analysis with different machine learning approaches on polycystic ovary syndrome," in *Proc. 23rd Int. Conf. Comput. Inf. Technol. (ICCIT)*, Dec. 2020, pp. 1–5.

[26] P. Mehrotra, J. Chatterjee, C. Chakraborty, B. Ghoshdastidar, and S. Ghoshdastidar, "Automated screening of polycystic ovary syndrome using machine learning techniques," in *Proc. Annu. IEEE India Conf.*, Dec. 2011, pp. 1–5.

[27] A. Denny, A. Raj, A. Ashok, C. M. Ram, and R. George, "I-HOPE: Detection and prediction system for polycystic ovary syndrome (PCOS) using machine learning techniques," in *Proc. IEEE Region 10 Conf. (TENCON)*, Oct. 2019, pp. 673–678.

[28] S. Bharati, P. Podder, and M. R. H. Mondal, "Diagnosis of polycystic ovary syndrome using machine learning algorithms," in *Proc. IEEE Region Symp. (TENSYMP)*, Jun. 2020, pp. 1486–1489.

[29] S. S. Deshpande and A. Wakankar, "Automated detection of polycystic ovarian syndrome using follicle recognition," in *Proc. IEEE Int. Conf. Adv. Commun., Control Comput. Technol.*, May 2014, pp. 1341–1346.

[30] M. Alagarsamy, N. Shanmugam, D. P. Mani, M. Thayumanavan, K. K. Sundari, and K. Suriyan, "Detection of polycystic syndrome in ovary using machine learning algorithm," *Int. J. Intell. Syst. Appl. Eng.*, vol. 11, no. 1, pp. 246–253, Jan. 2023.

[31] E. Nsugbe, "An artificial intelligence-based decision support system for early diagnosis of polycystic ovaries syndrome," *Healthcare Anal.*, vol. 3, Nov. 2023, Art. no. 100164.

[32] D. N. S. Hari, P. Vanaja, M. A. Kumar, M. D. V. S. Akash, and K. Sivaiah, "Multi disease detection using machine learning," *Int. J. Food Nutritional Sci.*, vol. 11, no. 12, pp. 1640–1650, 2022.

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

[34] R. R. Vasavi, S. P. Prathibha, H. Valiveti, S. Maringanti, and A. Parsa, "Polycystic ovary syndrome monitoring using machine learning," in *Proc. Int. Conf. Intell. Data Commun. Technol. Internet Things (IDCIoT)*, Jan. 2023, pp. 1013–1019.

[35] S. Aggarwal and K. Pandey, "Early identification of PCOS with commonly known diseases: Obesity, diabetes, high blood pressure and heart disease using machine learning techniques," *Exp. Syst. Appl.*, vol. 217, May 2023, Art. no. 119532.

[36] D. Hdaib, N. Almajali, H. Alquran, W. A. Mustafa, W. Al-Azzawi, and A. Alkhayyat, "Detection of polycystic ovary syndrome (PCOS) using machine learning algorithms," in *Proc. 5th Int. Conf. Eng. Technol. Appl. (IICETA)*, May 2022, pp. 532–536. [Online]. Available: https://www.researchgate.net/publication/363925044_Detection_of_Polycystic_Ovary_Syndrome_PCOS_Using_Machine_Learning

[37] H. D. Mehr and H. Polat. (Oct. 28, 2021). *Diagnosis of Polycystic Ovary Syndrome Through Different Machine Learning and Feature Selection Techniques*. Accessed: Nov. 24, 2022. [Online]. Available: https://link.springer.com/article/10.1007/s12553-021-00613-y

[38] I. Kyrou, E. Karteris, T. Robbins, K. Chatha, F. Drenos, and H. S. Randeva, "Polycystic ovary syndrome (PCOS) and COVID-19: An overlooked female patient population at potentially higher risk during the COVID-19 pandemic," *BMC Med.*, vol. 18, no. 1, p. 220, Dec. 2020.

[39] S. A. Bhat and R. Gupta. (2021). *National College of Ireland Project Submission Sheet School of Computing*. Accessed: May 16, 2022. [Online]. Available: http://norma.ncirl.ie/5137/1/shakoorahmadbhat.pdf

[40] M. S. K. Inan, R. E. Ulfath, F. I. Alam, F. K. Bappee, and R. Hasan, "Improved sampling and feature selection to support extreme gradient boosting for PCOS diagnosis," in *Proc. IEEE 11th Annu. Comput. Commun. Workshop Conf. (CCWC)*, Jan. 2021, pp. 1046–1050.

[41] S. Alshakrani, S. Hilal, and A. M. Zeki, "Hybrid machine learning algorithms for polycystic ovary syndrome detection," in *Proc. Int. Conf. Data Analytics Bus. Ind. (ICDABI)*, Oct. 2022, pp. 160–164.

[42] A. Gupta, H. Soni, R. Joshi, and R. M. Laban, "Discriminant analysis contrasting dimensions for polycystic ovary syndrome ognostication," 2022. [Online]. Available: https://arxiv.org/abs/2201.03029

[43] S. A. Suha and M. N. Islam, "Exploring the dominant features and data-driven detection of polycystic ovary syndrome through modified stacking ensemble machine learning technique," *Heliyon*, vol. 9, no. 3, p. e14518, 2023.

[44] E. Sterling. (Nov. 8, 2011). *Hormone Levels and PCOS*. Contemporary OB/GYN. Accessed: May 16, 2022. [Online]. Available: https://www.contemporaryobgyn.net/view/hormone-levels-and-pcos

[45] M. D. Kahsar-Miller and R. Azziz, "The effectiveness of the interview for predicting the presence of polycystic ovary syndrome," *Gynecological Endocrino.*, vol. 17, no. 6, pp. 449–454, Jan. 2003, doi: 10.1080/09513590312331290378.

[46] *Sciencedirect.com*. Accessed: May 16, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S1751721415001396

[47] I. El Naqa and M. J. Murphy, "What is machine learning?" in *Machine Learning in Radiation Oncology*. Cham, Switzerland: Springer, 2015, pp. 3–11.

[48] *Unstop—Competitions, Quizzes, Hackathons, Scholarships and Internships for Students and Corporates*. Accessed: May 16, 2022. [Online]. Available: https://unstop.com/blog/classification-vs-clustering

[49] V. Jakkula. *Tutorial on Support Vector Machine (SVM)*. Accessed: May 16, 2022. [Online]. Available: https://course.ccs.neu.edu/cs5100f11/resources/jakkula.pdf

[50] I. Rish, T. J. Watson, and R. Center. *An Empirical Study of the Naive Bayes Classifier*. Accessed: May 16, 2022. [Online]. Available: https://www.cc.gatech.edu/home/isbell/classes/reading/papers/Rish.pdf

[51] G. Guo, H. Wang, D. Bell, Y. Bi, and K. Greer, ''KNN model-based approach in classification,'' in *On The Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE*. Berlin, Germany: Springer, 2003, pp. 986–996.

[52] Y. Ben-Haim. *A Streaming Parallel Decision Tree Algorithm*. Accessed: May 16, 2022. [Online]. Available: https://www.jmlr.org/papers/volume11/ben-haim10a/ben-haim10a.pdf

[53] G. Biau and E. Scornet, ''A random forest guided tour,'' *TEST*, vol. 25, no. 2, pp. 197–227, Jun. 2016.

[54] *Sciencedirect.com*. Accessed: May 16, 2022. [Online]. Available: https://www.sciencedirect.com/topics/computer-science/logistic-regression#:~:text=Logistic%20regression%20is%20a%20process,%2Fno%2C%20and%20so%20on

[55] *K-Means Clustering Algorithm*. Accessed: May 16, 2022. [Online]. Available: www.javatpoint.com and https://www.javatpoint.com/k-means-clustering-algorithm-in-machine-learning

[56] Y.-C. Wu and J.-W. Feng, ''Development and application of artificial neural network,'' *Wireless Pers. Commun.*, vol. 102, pp. 1645–1656, Sep. 2017.

[57] S. Albawi, T. A. Mohammed, and S. Al-Zawi, ''Understanding of a convolutional neural network,'' in *Proc. Int. Conf. Eng. Technol. (ICET)*, Aug. 2017, pp. 1–6.

[58] C. Bento. (Sep. 21, 2021). *Multilayer Perceptron Explained With a Real-Life Example and Python Code: Sentiment Analysis*. Towards Data Science. Accessed: May 16, 2022. [Online]. Available: https://towardsdatascience.com/multilayer-perceptron-explained-with-a-real-life-example-and-python-code-sentiment-analysis-cb408ee93141

SAMIA AHMED received the B.Sc. degree in information and communication engineering from the Bangladesh University of Professionals (BUP), in 2022, where she is currently pursuing the M.Sc. degree in ICE. She is a Teaching Assistant with BUP. She is conducting her research in ML. Her research interest includes healthcare with ML. In 2023, she was a recipient of the Chancellor's Gold Medal.

MD. SAZZADUR RAHMAN received the B.Sc. and M.S. degrees in applied physics, electronics, and communication engineering from the University of Dhaka, Dhaka, Bangladesh, in 2005 and 2006, respectively, and the Ph.D. degree in material science from Kyushu University, Japan, in 2015. He was a Faculty Member with the Faculty of Computer Science and Engineering, Hajee Mohammad Danesh Science and Technology University, from May 2009 to November 2018. Since November 2018, he has been with the Institute of Information Technology, Jahangirnagar University, Savar, Dhaka, where he is currently an Associate Professor. His research interests include material science for computer applications, surface science, ubiquitous computing, WSNs, machine learning, and the IoT.

ISMATE JAHAN received the B.Sc. degree in information and communication engineering from the Bangladesh University of Professionals (BUP), in 2022, where she is currently pursuing the M.Sc. degree in ICE. Her research interest includes the uses of IoT in healthcare and medical systems.

M. SHAMIM KAISER (Senior Member, IEEE) received the bachelor's and master's degrees in applied physics, electronics, and communication engineering from the University of Dhaka, Bangladesh, in 2002 and 2004, respectively, and the Ph.D. degree in telecommunication engineering from the Asian Institute of Technology (AIT), Pathum Thani, Thailand, in 2010. In 2005, he joined the Department of ETE, Daffodil International University, as a Lecturer. In 2010, he was an Assistant Professor with the Department of EEE, Eastern University, Bangladesh, and the Department of MNS, BRAC University, Dhaka. Since 2011, he has been an Assistant Professor with the Institute of Information Technology, Jahangirnagar University, Dhaka, where he became an Associate Professor in 2015, and a Full Professor, in 2019. He is currently the Director of the Institute of Information Technology, Jahangirnagar University. He has authored more than 100 papers in different peer-reviewed journals and conferences. His research interests include data analytics, machine learning, wireless networks and signal processing, cognitive radio networks, big data and cyber security, and renewable energy. He is the Founding Chapter Chair of the IEEE Bangladesh Section Computer Society Chapter.

A. S. M. SANWAR HOSEN (Member, IEEE) received the M.S. and Ph.D. degrees in computer science and engineering from Jeonbuk National University, Jeonju, South Korea. He is currently an Assistant Professor with the Department of Artificial Intelligence and Big Data, Woosong University, Daejeon, South Korea. He has published several papers in journals and international conferences. His research interests include wireless sensor networks, the Internet of Things, fog-cloud computing, cyber security, artificial intelligence, and blockchain technology. He has been invited to serve as the Technical Program Committee Member in several reputed international conferences, such as IEEE ACM. He has been an expert reviewer for IEEE Transactions, Elsevier, Springer, and MDPI journals and magazines.

DEEPAK GHIMIRE received the bachelor's degree in computer engineering from Pokhara University, Nepal, and the M.S. and Ph.D. degrees in computer science and engineering from Jeonbuk National University, in 2011 and 2014, respectively. He was an Assistant Lecturer with Pokhara University. He was a Postdoctoral Researcher with the Intelligent Image Processing Research and Development Division, Korea Electronics Technology Institute, from 2014 to 2020, engaged in projects related to advanced driver assistance systems. He is currently a Research Professor with Soongsil University, South Korea, with expertise in deep learning, computer vision, and image processing. He has ben an AI Computer Vision Engineer with NVISO, a Swiss-based company specializing in AI Edge Computing applications, since 2018. He has published more than 35 research papers in international journals and conferences.

SEONG-HEUM KIM received the bachelor's degree in electrical and electronic engineering from Yonsei University, in 2008, and the master's and Ph.D. degrees in electrical engineering from KAIST University, in 2010 and 2018, respectively. During the graduate studies, he was with the Robotics and Computer Vision Laboratory, KAIST, under the supervision of Prof. In So Kweon and Yu-Wing Tai. He was with the LG Electronics' System IC Center, from 2011 to 2014, and the Korea Electronics Technology Institute's Intelligent Information Research and Development Division, from 2017 to 2020, gaining valuable industrial experience. He is currently an Assistant Professor with Soongsil University, where he focuses on deep learning, machine learning, and image processing. His research interests include 3D pose estimation, action recognition, autonomous driving, and lightweight deep learning.