A Review on Polycystic Ovary Syndrome(PCOS) Detection

**Mohammad Rafi1, Ranjitha J2,Rakshitha C3, Sneha G4,Tejaswini Surve K S5** [mdrafi2km@yahoo.com1,](mailto:mdrafi2km@yahoo.com1) [ranjitharanju5@gmail.com2,](mailto:ranjitharanju5@gmail.com2) [rakshithachidanand342@gmail.com3,](mailto:rakshithachidanand342@gmail.com3) snehag042001@gmail.com4,tejusurve162@gmail.com5

Department of Computer Science and Engineering, UBDTCE

# 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**[1].** It is the cause of various problems, including infertility. It is one of the most common causes of oligo-ovulatory infertility, affecting approximately 4% of unselected reproductive-aged women**[1] [2]**. 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, androgens, hirsutism, genetics, machine learning, convolutional neural network.

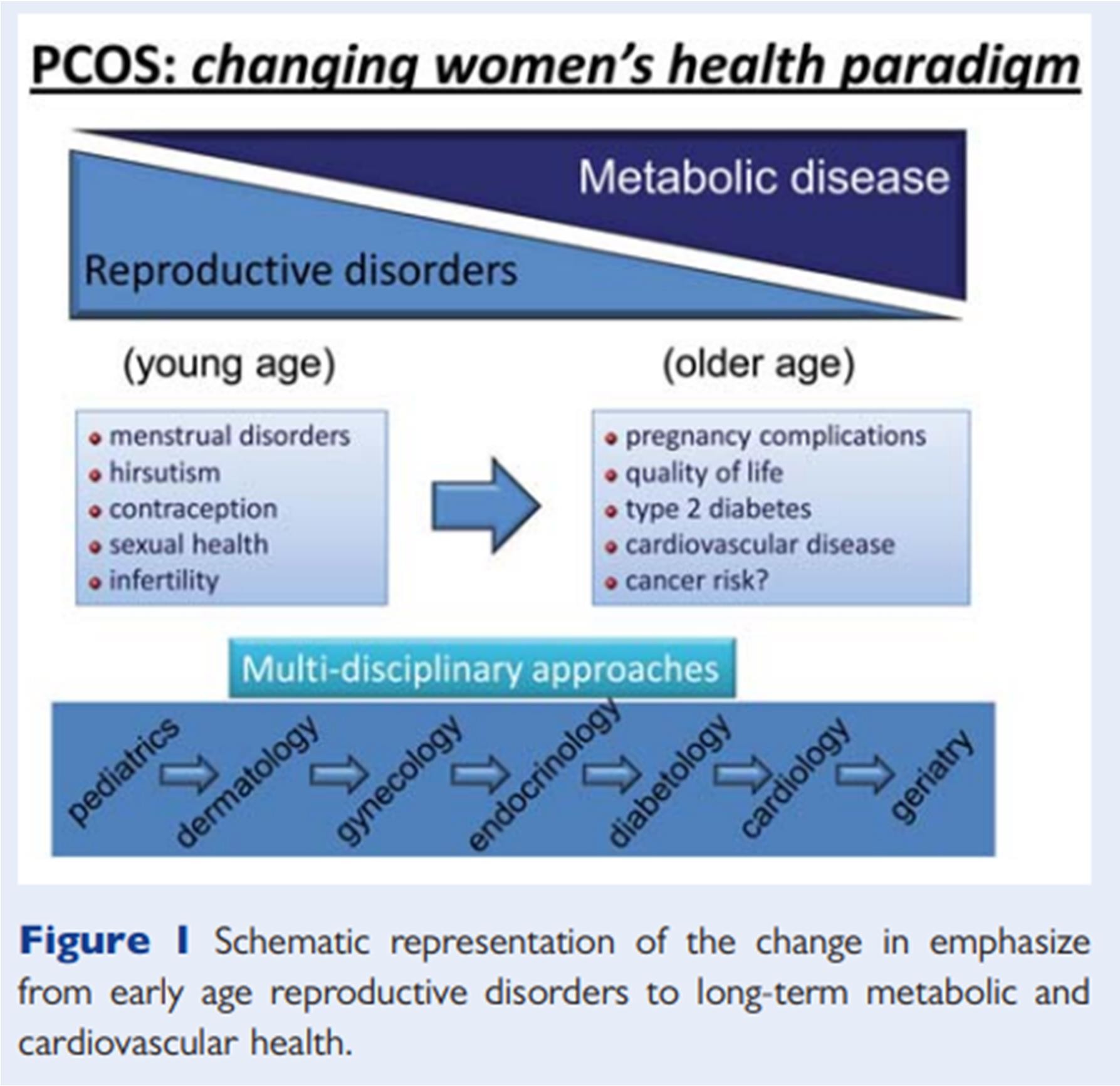


# INTRODUCTION

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 [3]. 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 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 acneare some of the symptoms [4]. 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 Schematic representation of the change in emphasize from early age reproductive disorders to long-term metabolic and cardiovascular health.

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 [6]. 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. 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 PCOS [5]. However, there are no such review articles regarding PCOS detection using ML that uses statistical parameters forcomparing 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.

# SYMPTOMS OF PCOS:

## Adolescence

There is no overall agreement as to how to diagnose PCOS in adolescence. Acne is common during the adolescent years, whether or not PCOS is present, whereas hirsutism— associated with PCOS—typically develops over time. Hyperandrogenemia may be a more consistent marker for PCOS during the teenage years (Blank et al., 2008). In all young women, irregular menses are common in the years immediately following menarche. As many as 85% of menstrual cycles are anovulatory during the first year after menarche, while up to 59% are still anovulatory during the third year following menarche (Apter, 1998). In one study, persisting oligomenorrhea was not predicted by increased androgens, polycystic ovaries on ultrasound or increased serum LH levels (van Hooff et al., 2004). Increased BMI, however, was the major risk factor for persistent anovulation. Only around 40%[11] of adolescent women with menstrual irregularity have polycystic ovaries on ultrasound (Venturoli et al., 1995). These considerations have led to the suggestion that all three elements of the Rotterdam criteria should be present in teenagers in order to make the diagnosis of PCOS (Carmina et al., 2010). These investigators suggest that oligomenorrhea or amenorrhea should be present for at least 2 years after menarch (or primary amenorrhea at age 16 years), the diagnosis of polycystic ovaries on ultrasound should include increased ovarian size (.10 cm3 ), and hyperandrogenemia rather than just signs of androgen excess should be documented.

## Hirsutism/acne/alopecia

Hirsutism is a good marker for hyperandrogenism even when considering ethnic differences and systemic factors such as obesity. Hirsutism is present in 70% of women with PCOS, but hyperandrogenemia should be evaluated biochemically in all women suspected of having PCOS. By comparison, acne and alopecia are not commonly associated with

hyperandrogenemia and therefore should not be regarded as evidence of hyperandrogenemia. For women with PCOS in whom hirsutism is a major concern, treatment is focused on reduction in androgen production, decreasing the fraction of circulating free testosterone (T) and limiting androgen bioactivity to hair follicles. In those women with PCOS who have acne vulgaris, clinical benefit may be derived from many systemic therapeutic modalities[3] [4]. Because terminal hair turnover occurs slowly, at least 6 months of treatment is generally considered the minimal interval to see a response.

## Menstrual irregularity

Although cycle abnormalities are common during the reproductive years, women with PCOS may ovulate spontaneously. How frequently this occurs is unknown (Laven et al., 2002), but ovulations have been reported in up to 32% of ‘cycles’. Women with oligo or amenorrhea have about a 90% chance of being diagnosed with PCOS and up to 95% of affected adults have oligo or amenorrhea (Kumarapeli et al., 2008). The definition used to establish the diagnosis of PCOS affects the proportion of women included with menstrual irregularities (Vutyavanich et al., 2007) [10]. Amenorrheic women with PCOS usually have the most severe hyperandrogenism and higher antral follicle counts when compared with women presenting with oligomenorrhea or regular menstrual cycles. Menstrual cycles in women with PCOS become more regular as they approach menopause (Dahlgren et al., 1992; Elting et al., 2001). One large study reported that obesity rather than the menstrual cycle pattern or the size of the follicular cohort determines hyperinsulinemia, dyslipidemia and hypertension in aging women with PCOS.

## Contraception

Women with PCOS who do not desire pregnancy need contraception. No contraceptive methods are contraindicated in PCOS. However, some of the features associated with PCOS [obesity, insulin resistance (IR) etc.] may represent a relative contraindication to the use of combined OCPs. Cycle control is usually achieved by the use of OCPs in women with PCOS. OCPs suppress LH secretion and lead to a decrease in ovarian androgen production. The estrogenic component increases the levels of SHBG, which, in turn, results in a decrease in circulating free T levels. The progestin in the pill can compete for 5a-reductase at the level of the androgen receptor. Oral contraception also decreases adrenal androgen production by a mechanism yet unclear, possibly due to a decrease in adrenocorticotropin hormone production.

## Quality of life

Patients with PCOS are an at-risk group for psychological and behavioral disorders and reduced quality-of-life (QoL) (Himelein and Thatcher, 2006; Jones et al., 2008; Dokras et al., 2011). Studies in this area have been hampered by the existence of only one validated diseasespecific questionnaire, the QoL questionnaire for women with PCOS (PCOSQ) (Cronin et al., 1998) [13]. A review of generic and specific QoL studies in women with PCOS concluded: (i) PCOS has a significant detrimental effect on QoL compared with controls,

(ii) weight issues were most apt to affect QoL, (iii) few studies included an instrument specific for PCOS in their assessment and (iv) very few studies included QoL instruments in their assessment of the benefits of the investigated treatment.

## Pregnancy

Women with PCOS may be subfertile. This may be explained

by the effects of obesity, metabolic, inflammatory and endocrine abnormalities on ovulatory function, oocyte quality and endometrial receptivity. Ovarian hyperandrogenism and hyperinsulinemia may promote premature granulosa cell luteinization and paracrine dysregulation of growth factors may disrupt the intrafollicular environment and impair cytoplasmic and/or nuclear maturation of oocytes (Dumesic et al., 2008). These features are not universal, and oocyte quality, fertilization and implantation rates in an individual woman with PCOS can be normal.

## Type 2 diabetes

IR is a prominent feature of PCOS. There is now compelling evidence from epidemiological data (Solomon et al., 2001) that PCOS is associated with increased risk of impaired glucose tolerance (IGT), GDM and T2D (Dunaif, 1997; Boomsma et al., 2006; Moran et al., 2010). Biochemical screening, in the form of an oral glucose tolerance test (OGTT), is indicated in obese women with PCOS, and/or those with increased visceral adiposity, as measured by waist circumference. Risk of IGT or diabetes is highest in women who have both oligo/anovulation and hyperandrogenism, and the risk is further amplified by obesity.

## Obesity

There is widespread variability in the prevalence of overweight (BMI 25– 30 kg/m2 ) and obese (BMI . 30 kg/m2 ) women in PCOS populations across different countries. The proportion of PCOS women who are overweight but not obese ranges from 10% in Italy to 37% in Kuwait. The highest prevalence of obesity is reported in studies conducted in USA and Australia, with 61 –76% of PCOS women considered obese.

# 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 [7] [12]. 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.

### 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[9]. 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) [14]. 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.

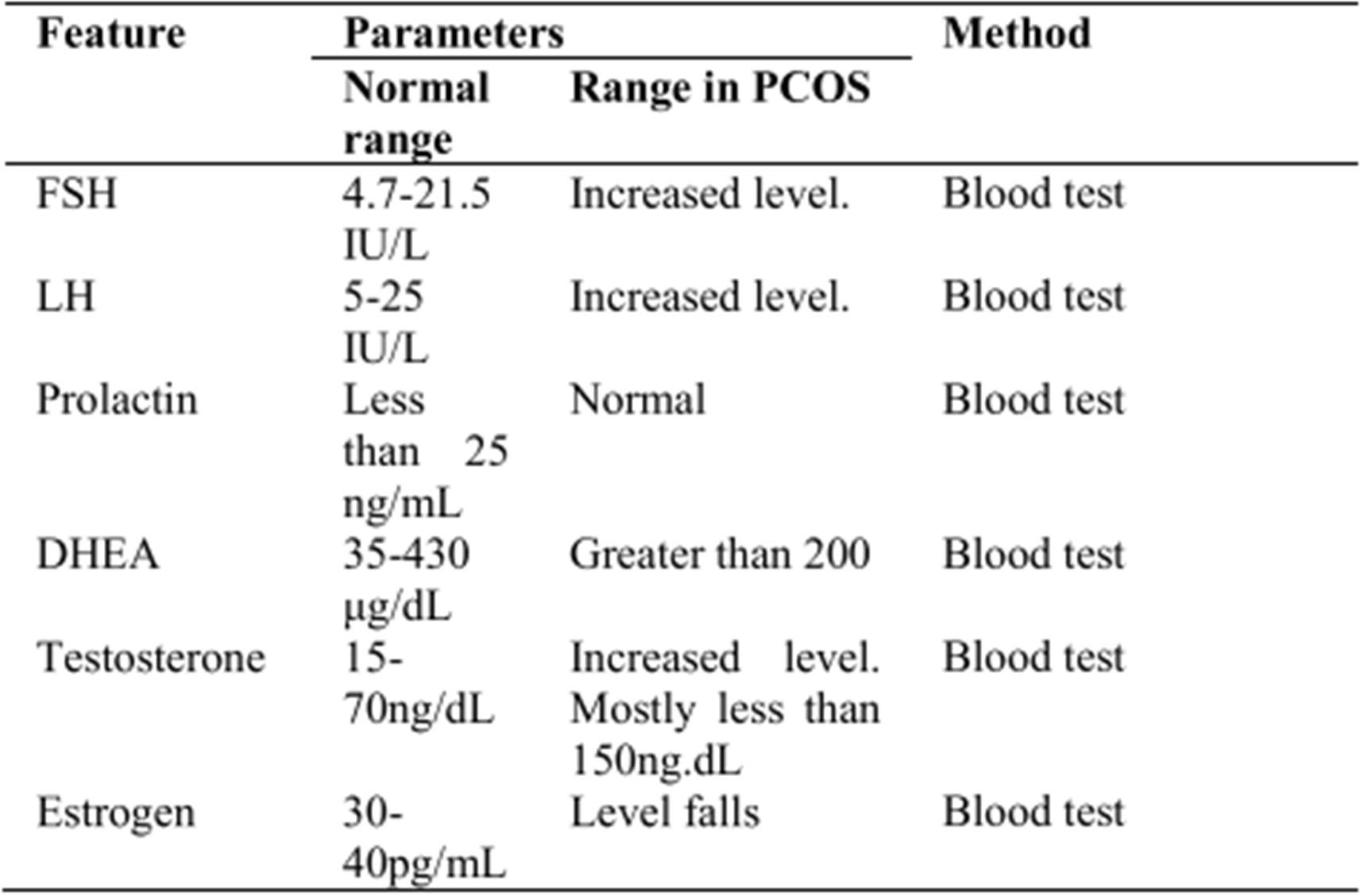


Fig: Normal range and the change of hormones in PCOS, along with the detected method.

1. **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[13] [14]. 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.

### MACHINE LEARNING (ML)

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

### 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[8]. 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. The classification algorithms are given below:

1. **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.

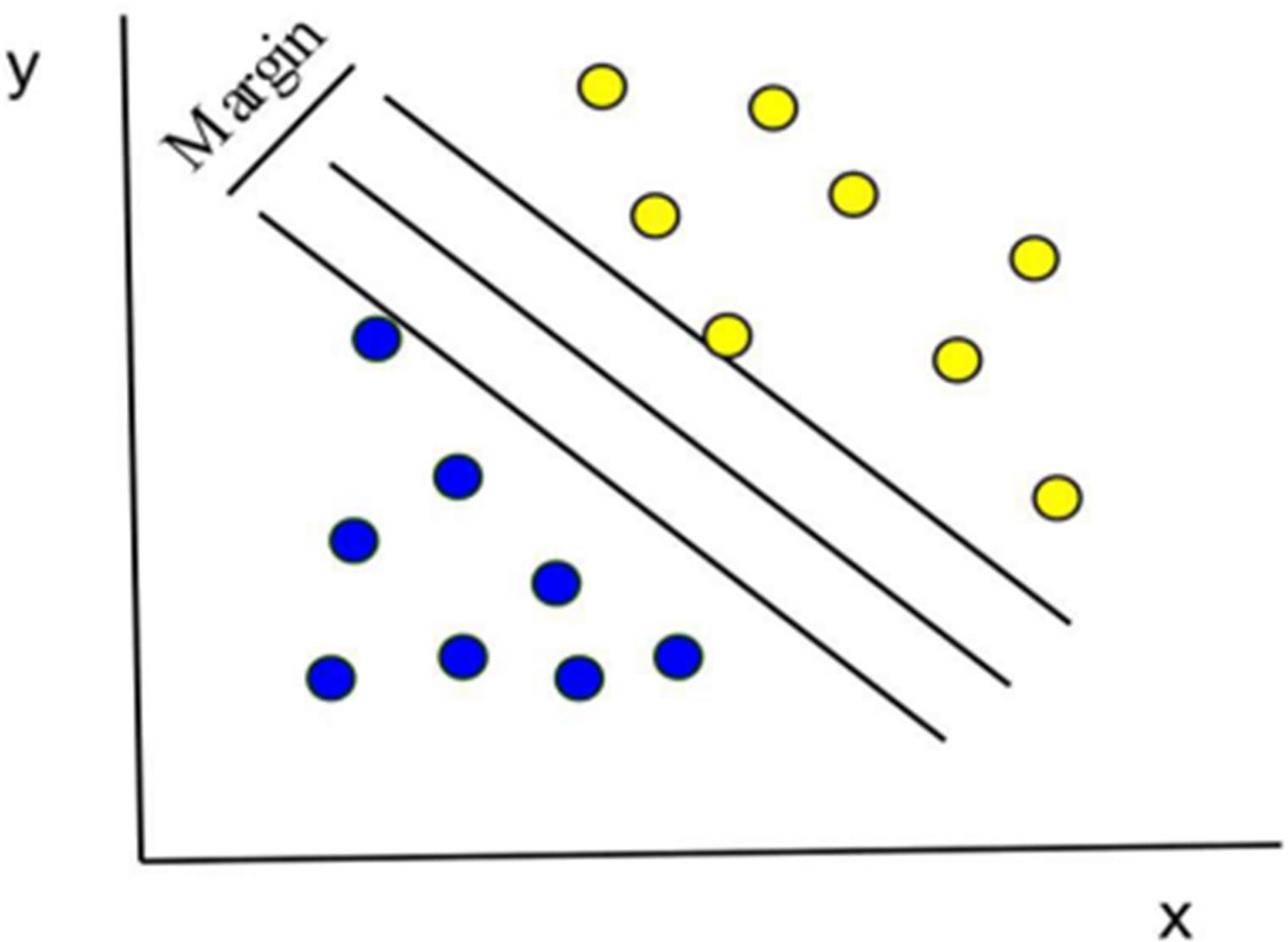


Fig: Visual representation of the working approach of SVM. Margin separates the two different classes generated by SVM.

Fig. 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.

1. **Naïve Bayes (NB):** It is a supervised learning algorithm based on Bayes theorem[15]. 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.
2. **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[16]. 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

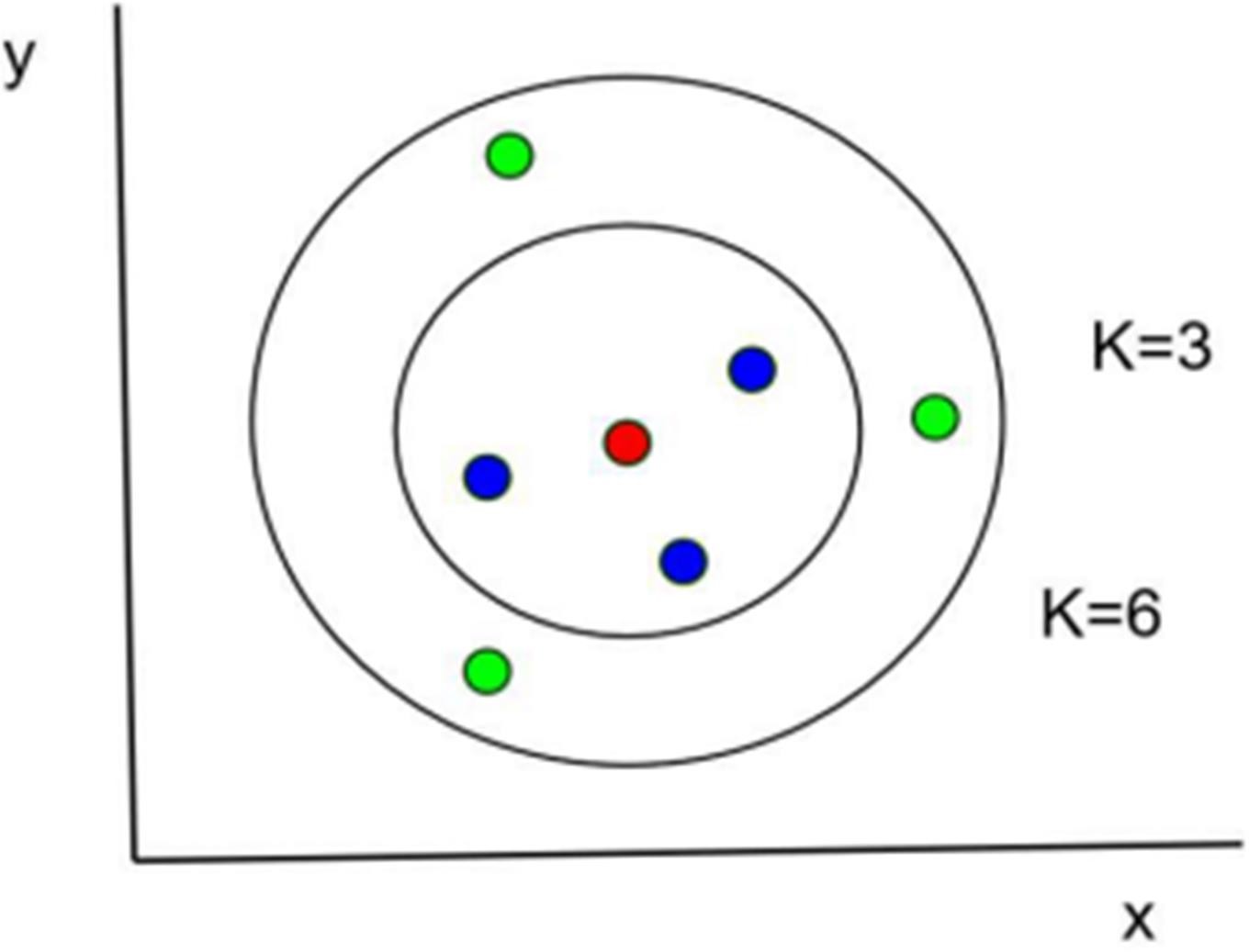


Fig: Representation of KNN. The red point is the origin. The blue points are classified when k=3, and when k=6, both the green and blue points are classified.

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 distancebased 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.

1. **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[14]-[16]. 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.

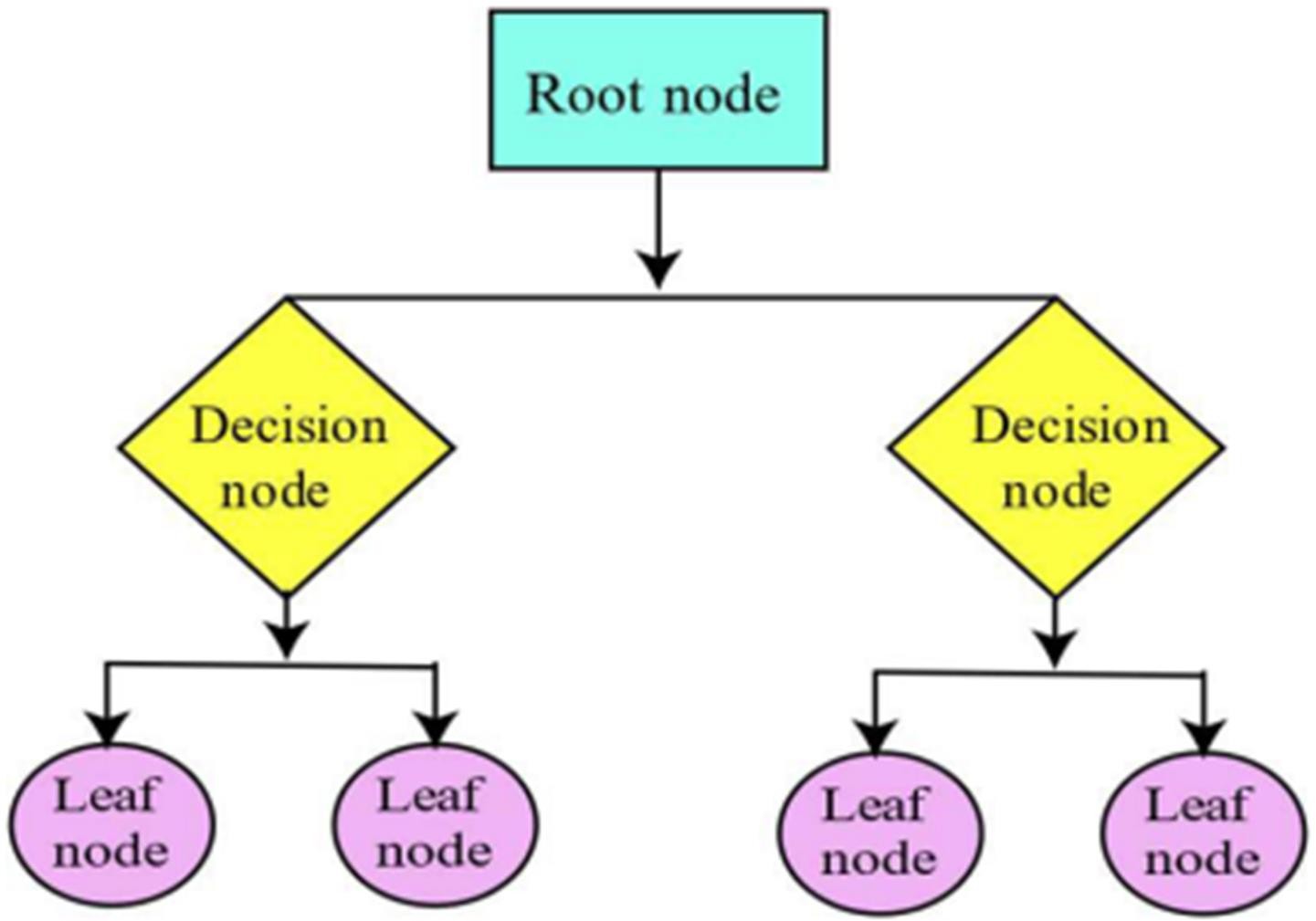


Fig: Working principle of a decision tree. The process starts at the root node. Decision nodes make leaf nodes and make various classes.

1. **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[17]. 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.

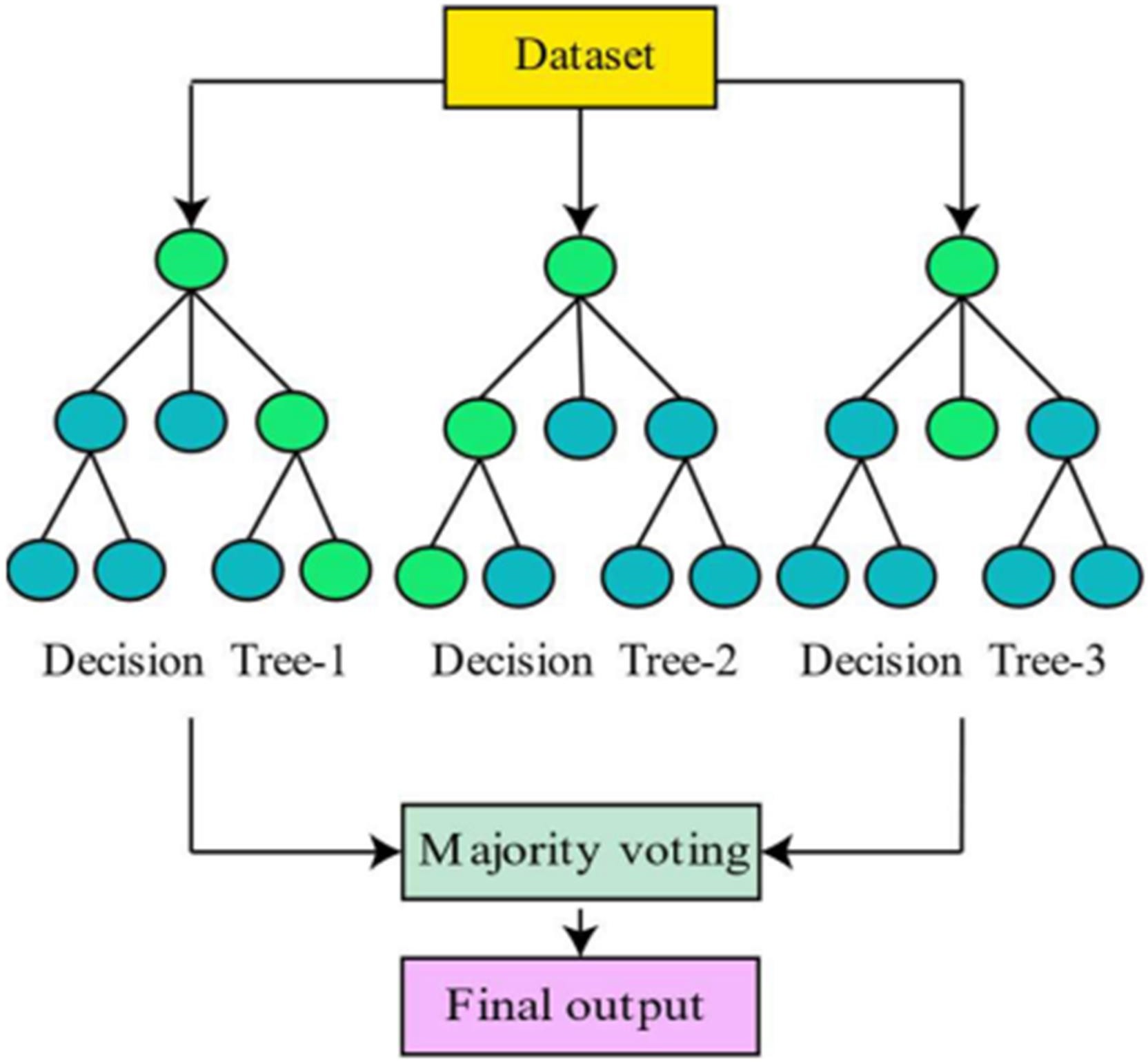


Fig: Visual presentation of random forest. Various decision trees are generated, and through majority voting, the final output is found.

1. **Logistic regression (LR):** It is a technique for estimating the likelihood of a binary outcome[19]. 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.

### 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 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[19] [20], 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. Thara [18] used adaptive k- means for PCOS detection. Fig. describes K-means. Readers can easily visualize the different clusters from this figure:

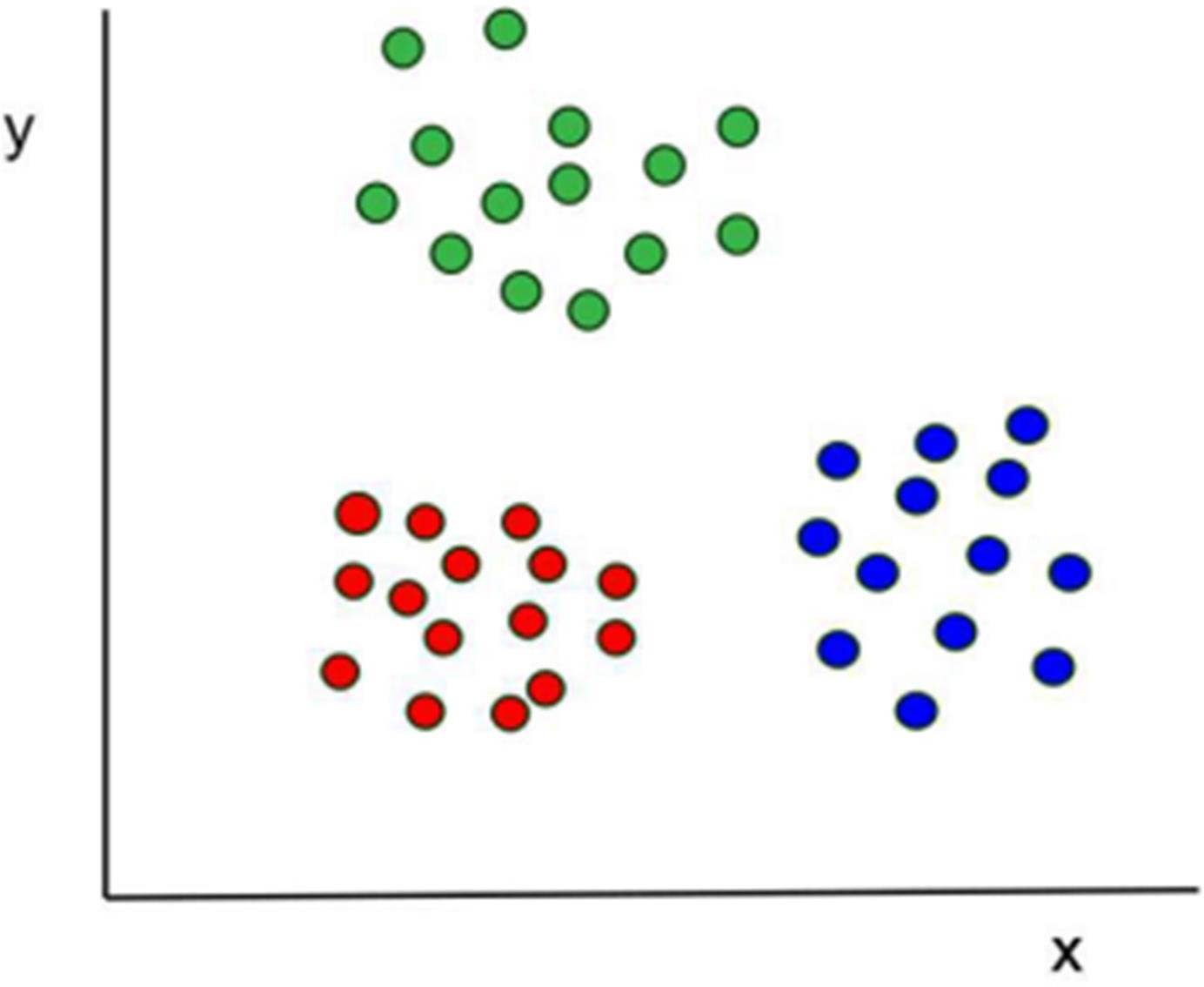


Fig: Visualization of the clusters generated by k-means algorithm.

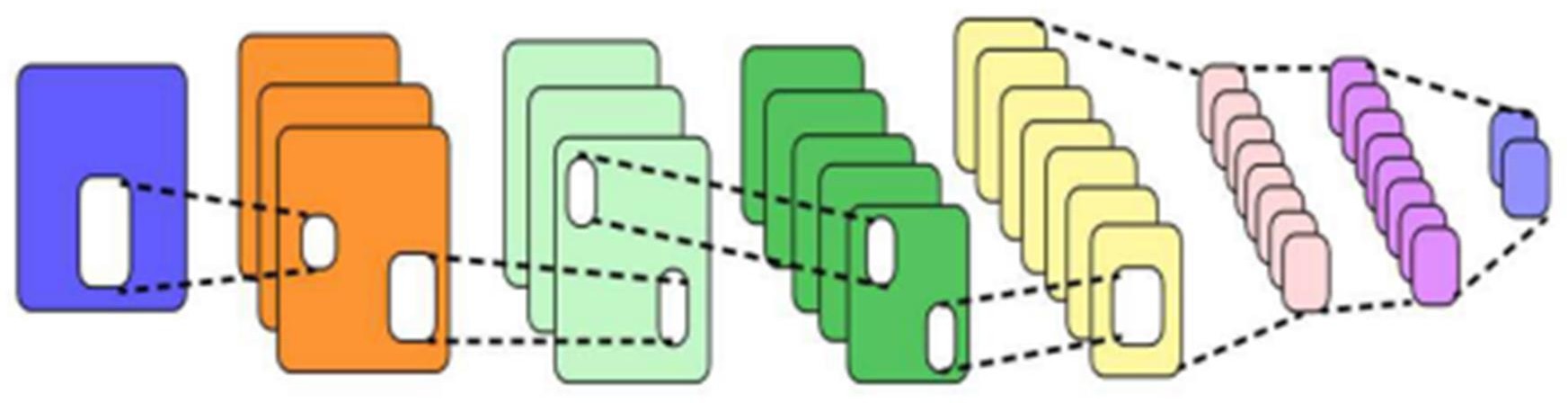


Fig: The architecture of CNN algorithm having various layers.

### 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[18]. 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].

* 1. **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 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

[24] 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. 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. Cahyonoetal [11] used CNN architecture to detect PCOS. This model achieved the perfect performance rate 0f 100%. Fig represents the architecture of CNN. The figure clearly represents various layers and their working principle of CNN.

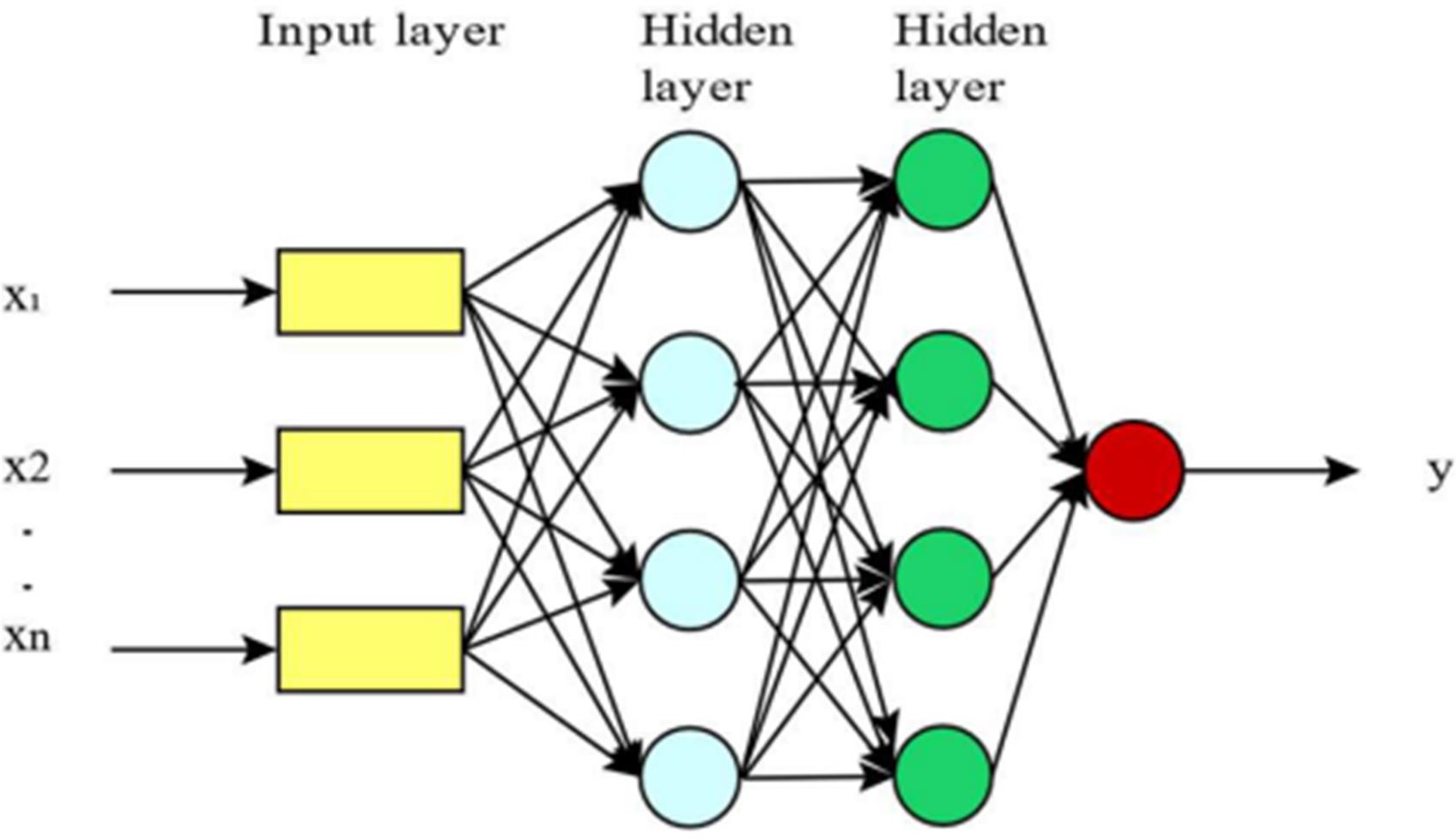


Fig: Vivid visualization of MLP architecture. The input layer is connected with multiple hidden layers.

* 1. **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 [25] [26]. It a very least used ANN technique for PCOS detection. Fig. depicts the architecture of MLP. This figure makes easy for researchers to understand the architecture and various layers of MLP.

# DATASET

This section gives a detailed overview of the datasets utilized in PCOS detection techniques discussed in this section. 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. Describes the dataset along with the source of the dataset. It will help researchers to choose the best dataset for future research work.

# 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 parameter help the further researchers to compare which model works better on that specific classification algorithm and which does not.

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[27]. 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. 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.

# VII. 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. 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.

1. **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[28]. 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.
2. **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 the other class has a low number of observations.
3. **NOISE IN ULTRASOUND IMAGE** To run CNN model, the images must be clear enough, but ultrasound images are prone to speckle noise[29], salt and pepper noise, and many other noises. These noises must be removed to get the perfect result.

# 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[29] [30]. 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-](http://www.nhs.uk/conditions/polycystic-ovary-syndrome-) pcos/

* 1. 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-](http://www.healthline.com/health/polycystic-ovary-) disease
  2. What’s the Treatment for PCOS? Accessed: May 16, 2022. [Online]. Available:

https://[www.webmd.com/women/treatment-pcos](http://www.webmd.com/women/treatment-pcos)

* 1. 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/339973](http://www.researchgate.net/publication/339973) 640\_Polycystic\_Ovary\_Syndrome\_Detection\_Using\_Vario us\_Machine\_Learning\_Methods-A\_Review
  2. Polson DW, Adams J, Wadsworth J, et al. Polycystic ovaries — a common finding in normal women. Lancet 1988; 1: 870-872.
  3. Knochenhauer ES, Key TJ, Kahsar-Miller M, et al. Prevalence of the polycystic ovary syndrome in unselected black and white women of the southeastern United States: a prospective study. J Clin Endocrinol Metab 1998; 83: 3078-3082.
  4. S. Bednarska, A. Siejka, The pathogenesis and treatment of polycystic ovary syndrome: what’s new?, Adv. Clin. Exp. Med. 26 (2017).
  5. V. Deepika, Applications of artificial intelligence techniques in polycystic ovarian syndrome diagnosis, J. Adv. Res. Technol. Manag. Sci. 1 (2019) 59–63.
  6. N. Thomas, A review on prognosis of pcos using nfrs,

hybrid technique and chi-square test, in: Proceedings of National Conference on Advanced Computing and Communication Technology, ISBN 97893-52886-869, 2019.

* 1. O.R. Isah, A. Usman, A. Tekanyi, A review on computer assisted follicle detection techniques and polycystic ovarian syndrome (pcos) diagnostic systems, 2015.
  2. T.L. Strome, Healthcare Analytics for Quality and Performance Improvement, John Wiley & Sons, 2013.
  3. N. Imtiaz Khan, T. Mahmud, M. Nazrul Islam, Covid-19 and black fungus: analysis of the public perceptions through machine learning, Eng. Rep. 4 (2022) e12475.
  4. K.T. Shahriar, M.N. Islam, M.M. Anwar, I.H. Sarker, Covid-19 analytics: towards the effect of vaccine brands through analyzing public sentiment of tweets, Inform. Med. Unlocked 31 (2022) 100969.
  5. M.I. Jordan, T.M. Mitchell, Machine learning: trends, perspectives, and prospects, Science 349 (2015) 255–260.
  6. M.N. Islam, K.R. Raiyan, S. Mitra, M.R. Mannan, T. Tasnim, A.O. Putul, A.B. Mandol, Predictis: an iot and machine learning-based system to predict risk level of cardio-vascular diseases, BMC Health Serv. Res. 23 (2023) 171.
  7. Hull MGR. Epidemiology of infertility and polycystic ovarian disease: endocrinological and demographic studies. Gynecol Endocrinol 1987; 1:235–45.
  8. Knochenhauer ES, Key TJ, Kahsar-Miller M, Waggoner W, Boots LR. Prevalence of the polycystic ovary syndrome in unselected black and white women of the southeastern United States: a prospective study. J Clin Endocrinol Metab 1998;83:3078–82.
  9. Case records of the Massachusetts General Hospital (case 43481). N Engl J Med 1957;108:257–62.
  10. Evans TN, Riley GM. Polycystic ovarian disease (Stein- Leventhal syndrome). Etiology and rationale for surgical treatment. Am J Obstet Gynecol 1958;12:168–78.
  11. Trace RJ, Keaty EC, McCall ML. An investigation of ovarian tissue and urinary 17-ketosteroids in patients with bilateral polycystic ovaries. Am J Obstet Gynecol 1960;79:310–5.
  12. Alsamarai S, Adams JM, Murphy MK, Post MD, Hayden DL, Hall JE, Welt CK. Criteria for polycystic ovarian morphology in polycystic ovary syndrome as a function of age. J Clin Endocrinol Metab 2009; 94:4961 – 4970.
  13. Apter D. Endocrine and metabolic abnormalities in adolescents with a PCOS-like condition: consequences for adult reproduction. Trends Endocrinol Metab 1998;9:58 – 61.
  14. 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.
  15. 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-areal-life-example-and-python-code- sentiment-analysis-cb408ee93141
  16. 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/pap](http://www.cc.gatech.edu/home/isbell/classes/reading/pap) ers/Rish.pdf

* 1. 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.
  2. 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
  3. 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/computerscience/logi](http://www.sciencedirect.com/topics/computerscience/logi) stic-regression#:~:text=Logistic%20regression%20is%20a

%20process,%2Fno%2C%20and%20so%20on

* 1. K-Means Clustering Algorithm. Accessed: May