

RESEARCH

Open Access



Optimized polycystic ovarian disease prognosis and classification using AI based computational approaches on multi-modality data

Kogilavani Shanmugavadivel¹, Murali Dhar M S², Mahesh T R³, Taher Al-Shehari⁴, Nasser A. Alsadhan⁵ and Temesgen Engida Yimer^{6*}

Abstract

Polycystic Ovarian Disease or Polycystic Ovary Syndrome (PCOS) is becoming increasingly communal among women, owing to poor lifestyle choices. According to the research conducted by National Institutes of Health, it has been observed that PCOS, an endocrine condition common in women of childbearing age, has become a significant contributing factor to infertility. Ovarian abnormalities brought on by PCOS carry a high risk of miscarriage, infertility, cardiac problems, diabetes, uterine cancer, etc. Ovarian cysts, obesity, menstrual irregularities, elevated amounts of male hormones, acne vulgaris, hair loss, and hirsutism are some of the symptoms of PCOS. It is not easy to determine PCOS because of its different combinations of symptoms in different women and various criteria needed for diagnosis. Taking biochemical tests and ovary scanning is a time-consuming process and the financial expenses have become a hardship to the patients. Thus, early prognosis of PCOS is crucial to avoid infertility. The goal of the proposed work is to analyse PCOS symptoms based on clinical data for early diagnosis and to classify into PCOS affected or not. To achieve this objective, clinical features dataset and ultrasound imaging dataset from Kaggle is utilized. Initially 541 instances of 45 clinical features such as testosterone, hirsutism, family history, BMI, fast food, menstrual disorder, risk etc. are considered and correlation-based feature extraction method is applied to this dataset which results in 17 features. The extracted features are applied to various machine learning algorithms such as Logistic Regression, Naïve Bayes and Support Vector Machine. The performance of each method is evaluated based on accuracy, precision, recall, F1-score and the result shows that among three models, Support Vector Machine model achieved high accuracy of 94.44%. In addition to this, 3856 ultrasound images are analysed by CNN based deep learning algorithm and VGG16 transfer learning algorithm. The performance of these models is evaluated using training accuracy, loss and validation accuracy, loss and the result depicts that VGG16 outperforms than CNN model with validation accuracy of 98.29%.

Keywords Polycystic ovary syndrome, VGG16, Transfer learning, Machine learning, CNN, Deep learning, Ultrasound images, Clinical features

*Correspondence:
Temesgen Engida Yimer
Temesgen.engida@du.edu.et
Full list of author information is available at the end of the article



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

Introduction

Five percent to ten percent of females between the ages of 18 and 44 suffer from polycystic ovarian syndrome (PCOS), a gynaecological endocrine disorder [1]. A menstrual cycle that is delayed or non-existent is the result of this hormonal imbalance. Many of the painful, unpleasant, and surprising symptoms of PCOS are also associated with unattractive and unfeminine culturally defined characteristics. In the ovaries, PCOS causes aberrant follicular growth that is prematurely terminated and never matures. This anomaly, which is the first indication of PCOS, is one of the reasons why people struggle to conceive. There is no single test that can be used to understand and diagnose PCOS. Instead, clinicians must rely on symptoms, blood tests and in certain cases, ultrasound scan, which provide information like the number of follicles and each follicle size to identify whether a person has polycystic ovarian syndrome. While the exact origins of PCOS are unknown, research shows that it is mostly hereditary. It is a very unpredictable ailment because there is no apparent pattern for this medical condition [2]. Both women and doctors struggle with the time and expense required for numerous medical tests and scans. A woman is reported to have PCOS if exhibits any one of the symptoms like increased androgen levels or polycystic ovaries, according to the Rotterdam guidelines or criteria [3]. According to the research conducted by MedlinePlus (<https://medlineplus.gov/genetics/condition/polycystic-ovary-syndrome/>), when one or both ovaries have 12 or more follicles, or when the ovary is larger than 12 cm, a woman is said to have polycystic ovaries and is therefore infected with this condition. To accurately forecast this problem, the doctor usually counts the number of follicles present in the ovaries from the ultrasound images in a manual manner. Furthermore, determining whether the women having PCOS or not is a time-consuming process and since this is sensitive issue, high level of accuracy is also expected.

Although there is a lot of evidence that PCOS has a role in disturbing reproductive health, there is little study on how to detect PCOS in women at an early stage. Unlike other disorders that may be predicted based on heredity, doctors are unsure who will develop PCOS and who will not. PCOS symptoms can be indicators, but just because you have them doesn't indicate you have PCOS. As a result, doctors are prone to overlooking PCOS. If a woman has excessive acne and visits a dermatologist, she is unlikely to inquire about other issues such as missed periods. Seeing a doctor, such as an OBGYN, who will understand all your symptoms and make the right diagnosis, is the best method to acquire a diagnosis. Blood tests and even an ultrasound to look for ovarian cysts are commonly used to diagnose PCOS. PCOS can become

a major problem if left untreated. All the symptoms that women experience can lead to cancer, acne scars, and heart disease, among other things. Sleep apnea and infertility are two other health issues to consider. PCOS has several dangers linked with it. Therefore, it is crucial to check individuals early to limit any serious effects of the disease. For early-stage PCOS identification and prediction, the currently available techniques and treatments are insufficient. Early detection and treatment are essential because PCOS has additional side effects that can be avoided with small lifestyle changes, such as type 2 diabetes or cardiovascular issues. This PCOS problem among women leads to early pregnancy miscarriages, struggle with infertility, and in rare cases, develop gynaecological cancer, this early detection also lessens the risks connected with the condition.

PCOS can be identified from ovarian ultrasound pictures using a machine learning model called PCONet that employs a Convolutional Neural Network (CNN) [4]. They also fine-tuned a pre-trained model called InceptionV3 for the same task using transfer learning. The performance evaluation shows that PCONet had 98.12% accuracy. An application is developed by Purnama et al. to classify PCOS by identifying follicles in ultrasound images [5]. Applying low pass filters, equating the histogram, binarizing the image, and employing morphological techniques to produce binary follicular images are all steps in the preparation stage. Using edge detection, labelling, and cropping are all steps in the segmentation process. The feature extraction stage uses Gabor wavelets to extract texture information from the cropped images, resulting in the production of two datasets for classification.

Three alternative techniques are used in the classification stage: a neural network, KNN and SVM classifiers using an RBF kernel in [6]. Among all the models, SVM model on C=40 achieved the best accuracy of 82.55% on one dataset and 78.81% on the other. Further a system was developed by Dewi et al. [7] for detecting PCO in women's reproductive systems using feature extraction and a Competitive Neural Network (CNN). Currently, a gynecologist must perform PCO detection manually, which takes longer and requires a high degree of accuracy. The system created in this research extracts features from ultrasound images using the Gabor Wavelet approach, and then it uses a CNN to classify the images according to predetermined attributes. A system test yielded an accuracy of 80.84% and a CPU time of 60.64 s for the CNN.

From the above observations, either clinical features or ultrasound images alone do not provide accurate classification of PCOS. The objective of this research work is to detect whether the patient is affected or not by utilizing

the models trained by both clinical features and ultrasound images. The proposed research work utilizes various machine learning models for analysing clinical data. Deep learning approaches as mentioned in [8, 9] known as Convolutional Neural Network (CNN) based image classification and VGG-16 pre-trained transfer learning model are used to analyse ultrasound images further. These approaches will be utilised to shorten the time it takes to predict PCOS with improved accuracy, reducing the risk of deadly consequences that can occur when diagnosis is delayed.

The ovaries, which are positioned in the uterus, are an important element of the female reproductive system. Women have two ovaries, which produce eggs and emit oestrogen and progesterone hormones. An egg matures in a pod called a follicle during a woman's menstrual cycle and it can be seen within the ovaries. This follicle or pod normally splits open and then discharge an egg. However, if the process is not done properly, the fluid which is available in the ovary can develop a cyst, resulting as polycystic ovaries. Many women may develop Polycystic Ovary Syndrome because of this hormonal imbalance. According to a study by the PCOS Society, one in ten women in India have polycystic ovarian syndrome (PCOS), a common endocrine system among women. Six adolescent girls are diagnosed with PCOS for every 10 women who have it. In 1935, PCOS was first identified. But in India, the problem continues to be poorly understood in general and usually stays unreported for years. This condition is estimated to afflict roughly 10 million women worldwide. The prevalence of PCOS is estimated in a range between 3 and 10%, but it is unknown which subpopulations, based on place of residence and race/ethnicity, are affected. According to a study conducted by the AIIMS endocrinology and metabolism department, 20–25% of Indian women of reproductive age have PCOS issues. 60% of PCOS-afflicted females are overweight, but 35–50% also has fatty livers. According to studies, PCOS was discovered to be common in 9.13% of women in South India and 22.5% of women in Maharashtra. Many aspects of the illness are still unclear due to the considerable variation in its symptoms and severity.

In general, 5–10% of women worldwide is suffered by PCOS which is a prevalent endocrine condition and a primary cause of persistent menstrual disorders and infertility [10]. Women with PCOS have high quantities of male hormones and low levels of female hormones, causing their menstrual cycle to vary. The ovaries enlarge with PCOS, and there are often many small cyst formations termed immature follicles. Menstrual irregularities, symptoms of hyperandrogenism such as acne, hirsutism, hair loss, and infertility are all signs of PCOS. PCOS has also been leads other persistent health issues like heart

disease, obesity, infertility, uterine cancer, and diabetes. Studies show that most women will experience at least one cyst at some point in their lives. Cysts are frequently silent, which makes it challenging to diagnose because they don't show any symptoms. Although the precise aetiology of PCOS is unknown, it is believed that several factors can affect it. Hormonal abnormalities such as high levels of androgens, Luteinizing Hormone (LH), and normal or repressed Follicle Stimulating Hormone (FSH) are the main contributors to the unbalanced LH/FSH ratio (FSH). The clinical signs of hyperandrogenism are also associated with insulin resistance and hyperinsulinemia. It is unclear what factors put women at risk for developing PCOS. However, it has been noted that in certain cases, the condition may have a genetic basis and that many lifestyle factors, including obesity, increase the chance of developing PCOS and hyperinsulinemia [11]. Numerous published studies look into the prevalence and common clinical parameters of PCOS in various regions. But none that explain the link between some of these factors and PCOS. Some researchers have tried to comprehend the factors that contribute to the development of PCOS and to estimate the risk associated with it to recognise and track the symptoms of PCOS at an earlier stage and to prevent further problems. PCOS is difficult to diagnose because of the many symptoms, gynaecological, clinical, and metabolic markers involved [12]. Patients with PCOS are now burdened by the length of time needed for various clinical tests and ovarian scans, as well as the associated expenses. This is a major factor in why women initially disregard the signs of PCOS and later experience its effects. Not everyone can afford to get tested or scanned [13].

It discovered patterns in collected clinical data identifying individuals with Metabolic Syndrome using machine learning technologies [14]. This method was based on attribution rules, a type of rule that is easy to understand for medical professionals who aren't specialists in statistics or meta-analysis. Ningyi Zhang et al. presented DeepGP, deep learning technique, convolutional neural networks and graph convolutional networks, with the purpose of discovering susceptible genes in five endocrine illnesses, including POS. Ten cross-validations were performed on an integrated reported dataset to calculate the efficiency of the method [15].

Tan J et al. stated that the presence of PCOS may have an impact on the participants' mental health, which could be linked to psychiatric difficulties among such patients because of their inability to conceive [16]. Depression and anxiety were found to be present in 27.5% and 13.3% of PCOS patients, respectively, compared to 3.0% and 2.0% in control subjects. The research stated that existing research on numerous environmental factors suggests

that the various ecological toxins, diet plans, food habits, and geographical location may play an important role in deteriorating reproductive health [17].

The joint use of myo-inositol and d-chiro-inositol therapy on PCOS patients was studied and the authors found that there is a decline in body weight, a rise in blood SHBG, and changes in the levels of FSH, LH, and insulin. Furthermore, after 6 months of treatment, the serum glucose level in the OGTT was reduced [18]. To assess the PCOS problem and quality of life in Pakistan, Around 500 Women of reproductive age who visited the gynaecology in Islamabad were studied in detail for PCOS. A checklist was created to find symptoms, such as problems, illogical or under-prescription medication [19–21].

Using a stacking ensemble machine learning technique, Suha et al. identified the photos as either PCOS or non-PCOS. A bagging or boosting ensemble model was employed as a meta-learner, and conventional models served as base learners [22]. To extract characteristics from the photos, a Convolutional Neural Network was created by combining many approaches using transfer learning [23]. They found that using the trained VGG16 technique as a feature extractor and the XGBoost model as an image classifier yielded the best results, with a 99.89% classification accuracy.

P. Chauhan et al. stated that machine learning techniques should be used to create an application for early PCOS prediction. Using Google Colab and Python, the required dataset is produced and refined by conducting a survey [24, 25]. The Gini coefficient is used to determine the trait's relevance. Several machine learning approaches are used for classification process. According to ChaobaKshetrimayum et al., PCOS can be identified based on and changes in the target tissue separation process during foetal expansion, metabolic disorders, prenatal and postnatal coverage as well as lifestyle and dietary factors later in life [26].

Sun et al. discussed that acupuncture with metformin can be used to treat PCOS problem in [27]. In that clinical trials are screened by performing meta-analysis. Ultrasonography parameters are utilized by Gyliene et al. in [28]. A review of recent advancements to detect PCOS based on symptoms was discussed in [29]. Combination of metformin with pioglitazone was used to treat PCOS problem was analysed in [30]. Details of applying pharmacotherapy and clinical features to diagnose PCOS are discussed in [31]. An observational study about various prediction models for PCOS detection are discussed in [32]. A population-based cohort study was conducted by Linda et al. in [33].

Using a stacking ensemble machine learning technique, Suha et al. identified the photos as either PCOS or non-PCOS. A bagging or boosting ensemble model

was employed as a meta-learner, and conventional models served as base learners [22]. To extract characteristics from the photos, a Convolutional Neural Network was created by combining many approaches using transfer learning [23]. They found that using the trained VGG16 architectures for feature extractor and the XGBoost model as an image classifier yielded the best results, with a 99.89% classification accuracy. Compared to other machine learning methods already in use, this strategy greatly increased accuracy while shortening training execution time. Currently, there is no definitive objective test for diagnosing PCOS, and it can be difficult to diagnose based on ultrasound images alone, as it requires manual tracing and measurement of follicles. This issue is addressed using machine learning techniques to automatically diagnose PCOS from ultrasound images. The proposed method uses a KNN classifier and produces an accuracy of 97% in classification. This could potentially save time, improved PCOS diagnoses accuracy and reduced the complications that can result from delayed diagnosis. The following Table 1 represents existing approaches using machine learning and deep learning models for PCOS detection. Using a stacking ensemble machine learning technique, Suha et al. identified the photos as either PCOS or non-PCOS. A bagging or boosting ensemble model was employed as a meta-learner, and conventional models served as base learners [22]. To extract characteristics from the photos, a Convolutional Neural Network was created by combining many approaches using transfer learning [23]. They found that using the pre-trained VGG16 for feature extraction and the XGBoost model as an image classifier yielded the best results, with a 99.89% classification accuracy.

Despite its widespread occurrence, diagnosing PCOS remains a complex process due to several key challenges. Firstly, the syndrome presents a wide range of symptoms, which vary significantly from one individual to another. While some women exhibit pronounced symptoms like irregular menstruation and excessive androgen levels, others may have more subtle manifestations, complicating timely detection. The overlap of PCOS symptoms with other conditions, such as hypothyroidism and adrenal hyperplasia, further contributes to diagnostic ambiguity.

Table 1 PCOS clinical dataset statistics

Clinical Dataset	PCOS Infected	PCOS Not infected	Total
Training Data	142	291	433
Test Data	35	73	108
Total	177	364	541

In addition, there is no single test for PCOS. Diagnosing the condition typically requires a combination of clinical evaluations, hormone level assessments, and ultrasound imaging to identify the number of ovarian follicles or enlarged ovaries. This multifaceted approach is often time-consuming, expensive, and can vary in accuracy depending on the clinician's expertise. The manual counting of ovarian follicles using ultrasound, for instance, can be inconsistent and labor-intensive, introducing variability in diagnosis.

Another significant challenge lies in the limited awareness and understanding of PCOS among both patients and healthcare providers. Many women with early-stage PCOS do not seek medical advice until they face issues related to infertility, by which point the condition may have progressed. Misdiagnosis or delayed diagnosis can also occur, as some healthcare providers may focus on isolated symptoms rather than recognizing the broader syndrome.

Given these diagnostic hurdles, there is a pressing need for improved methods that can provide early, accurate, and cost-effective diagnosis of PCOS. Addressing these challenges is crucial for reducing the risk of complications such as infertility, metabolic syndrome, and cardiovascular diseases, which are often associated with untreated PCOS.

Technological advancements

Technological advancements in healthcare have significantly improved the early detection and prognosis of Polycystic Ovary Syndrome (PCOS), enabling more accurate and timely diagnoses. One of the key breakthroughs is the integration of artificial intelligence (AI) and machine learning (ML) into medical diagnostics. These technologies allow for the analysis of large, complex datasets, identifying patterns and correlations that are often missed through traditional diagnostic methods. Machine learning algorithms, for instance, can process clinical data, such as hormone levels and ultrasound images, to predict the likelihood of PCOS, even in its early stages.

Deep learning, a subset of machine learning, has further advanced the capabilities of diagnostic tools, particularly in image analysis. Convolutional Neural Networks (CNNs), a type of deep learning architecture, have proven highly effective in analyzing medical images, such as ultrasounds used in detecting ovarian abnormalities in women with PCOS. Pre-trained models like VGG-16, which utilize transfer learning, are now employed to process ultrasound images, enabling faster and more accurate identification of PCOS-related patterns.

Another significant technological advancement is the development of wearable health devices and Internet of

Medical Things (IoMT). These devices allow continuous monitoring of physiological parameters, such as blood sugar levels, heart rate, and physical activity, which are crucial for managing and detecting conditions associated with PCOS, such as insulin resistance and metabolic syndrome. The data collected from these devices can be analyzed in real-time, providing early warnings and personalized insights into a patient's health.

Additionally, advancements in telemedicine and cloud-based healthcare platforms enable remote diagnostics and monitoring. With the help of 5G technology, high-resolution imaging and clinical data can be transmitted and analyzed by healthcare professionals without requiring in-person visits. This is particularly useful for patients in remote or underserved areas, ensuring early detection and treatment of PCOS-related complications. In combination, these technologies are transforming the landscape of PCOS detection and prognosis, enabling more accurate, efficient, and accessible healthcare solutions for women affected by this condition.

Materials and methods

In healthcare systems, artificial intelligence techniques can be utilised to manage massive volumes of clinical data with great accuracy and precision. To predict PCOS problem, the proposed system first collects multi-modality data such as clinical features dataset and ultrasound images dataset from Kaggle. For the clinical features dataset correlation-based feature extraction is applied and important features are extracted. These extracted features are applied to various machine learning models like Logistic Regression (LR), Naïve Bayes (NB) and Support Vector Machine (SVM).

The performance of each machine learning model depends on the nature of the dataset and the classification task. Logistic Regression, being a linear model, performs well in cases where the relationship between the features and the target variable is mostly linear. However, since PCOS is a complex syndrome with nonlinear relationships among the clinical features, Logistic Regression may have limitations in capturing those complexities, which could result in lower predictive accuracy compared to more sophisticated models. Naïve Bayes, on the other hand, assumes feature independence and applies Bayes' theorem for classification. While it works efficiently with smaller datasets and is computationally faster, the assumption of independence between features can limit its performance in a dataset where features like hormone levels, BMI, and family history are interrelated. This likely leads to moderate performance in the context of PCOS prediction.

Support Vector Machine (SVM) often outperforms other models in handling nonlinear relationships,

especially when used with appropriate kernel functions. It is highly effective in high-dimensional spaces, which is advantageous when dealing with complex medical datasets. SVM's ability to create optimal decision boundaries and its robustness in separating data classes makes it suitable for the PCOS dataset. As a result, SVM typically achieves higher accuracy, precision, and recall compared to LR and NB models in this context. The superior performance of SVM can be attributed to its capacity to generalize well even with a relatively small dataset, which is common in medical applications. Thus, while Logistic Regression and Naïve Bayes provide valuable baseline comparisons, the nonlinear nature of PCOS and the interdependencies of clinical features favor more advanced models like SVM, leading to its higher accuracy in predicting PCOS.

For the ultrasound images dataset, Convolutional Neural Network based deep learning and VGG16 transfer learning models are applied. The proposed methodology workflow is specified in Fig. 1. The steps involved are data collection, data pre-processing, feature selection for text data, augmentation for image data, modelling with

machine learning, deep learning and transfer learning algorithms, model evaluation and prediction.

Dataset description

Two types of datasets such as PCOS clinical dataset which contains all physical and clinical parameters of patients, and PCOS ultrasound images dataset are used for this research work. Both the datasets are collected from Kaggle in which PCOS clinical dataset contains 541 records of PCOS and non PCOS information that spans over 45 columns [34]. PCOS ultrasound images dataset contains 1924 training set images and 1932 test set images. The statistics about the PCOS clinical dataset description is presented in Table 1.

In addition to clinical factors like blood group, pulse rate, HB, luteinizing hormone, human chorionic gonadotropin, number of follicles, thyroid stimulating hormone, AMH, PRL, and PRG, the suggested study also considers general physical features like age, weight, height, and BMI. To assess whether a woman had PCOS, it also included several other symptoms, such as weight gain, excessive hair growth, darkening of the skin, hair loss, and acne. The proposed system also considers PCOS

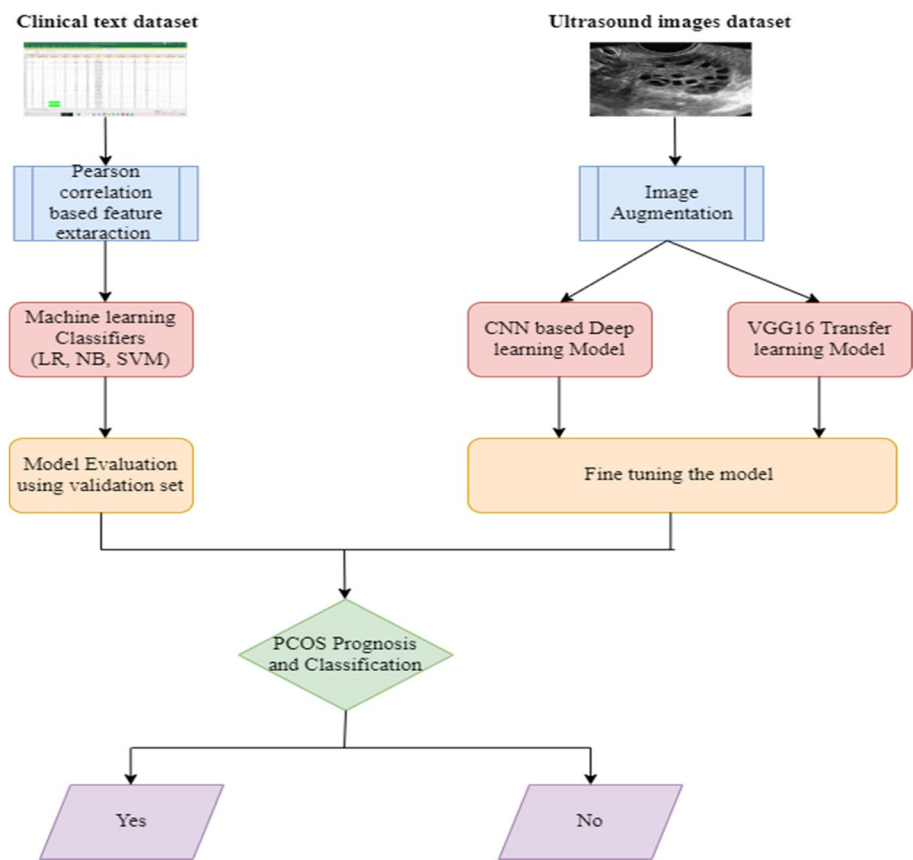


Fig. 1 Proposed system workflow

ultrasound images dataset that have been collected from Kaggle [35]. The detailed dataset description is depicted in Table 2.

Geometric metrics such as area, perimeter, compactness, smoothness, and accuracy along with the number of follicles are used to diagnose ultrasound images in detail. The Rotterdam criterion is one of the best methods for spotting PCOS. If the ovarian volume is greater than or equal to 10 cm or if there are more than 10 follicles with a diameter of 2 to 9 mm each, PCOS can be identified using ultrasound imaging. Figure 2 shows a single ovary that is unaffected, and several follicles affected by PCOS that are scattered across the ovary's periphery.

Feature extraction for clinical dataset

Feature extraction is the process of selecting important features from a larger set of attributes to improve the performance of a machine learning techniques Prior to creating the model, this stage of data preprocessing is crucial. Feature extraction aims to identify the most relevant and helpful features that will significantly affect the performance. The suggested techniques for clinical dataset features makes use of a Pearson correlation-based feature selection technique to choose characteristics with a high correlation with the target variable and a low correlation with other features. By calculating the correlation coefficient between each characteristic and the variable, it is possible to determine which features have the highest correlation with the target variable. The next stage is to identify the characteristics' relevance after the pertinent features have been chosen. Finding each feature in the dataset's relative relevance based on how much it

contributes to the target variable's prediction is known as feature importance identification. The size of each feature's coefficient is examined using the suggested system to assess the significance of each feature.

The purpose of feature extraction in the clinical dataset is to identify and select the most relevant variables that significantly contribute to the prediction of PCOS, thus improving the model's performance and reducing computational complexity. In the proposed system, feature extraction is carried out using a correlation-based method. This involves analyzing the relationships between various clinical features, such as testosterone levels, BMI, and menstrual irregularities, to determine their relevance to PCOS. The method filters out less relevant features, resulting in a reduced set of 17 key features that are most strongly correlated with the condition. These selected features are then used to train machine learning models, such as Logistic Regression, Naïve Bayes, and Support Vector Machines, enhancing the model's ability to accurately classify and predict PCOS based on clinical data.

Image augmentation for ultrasound images dataset

To train a machine learning model, its parameters are changed to fit a particular input (an image) and output (a label). The main objective is to build an efficient model with the least amount of model loss, but this is only possible when the parameters are specified properly. When there are many parameters, the machine learning model needs more examples to work well. The number of parameters needed also indicates how challenging the task is for the machine learning model to complete. Things could occasionally go wrong even with the right size training set. It's crucial to keep in mind that while humans may use their natural skills to categorize photos, algorithms do not have human-like intelligence. If a user builds a model that can discriminate between cats and dogs, but almost all the training images of dogs are of snowy landscapes, the algorithm can end up learning the wrong rules. Therefore, it's crucial to capture photos in a variety of circumstances and from a variety of angles. To get more data without adding extra training data, the

Table 2 PCOS ultrasound images dataset statistics

Ultrasound Images Dataset	PCOS Infected	PCOS Not infected	Total
Training Data	781	1143	1924
Test Data	787	1145	1932
Total	1568	2288	3856

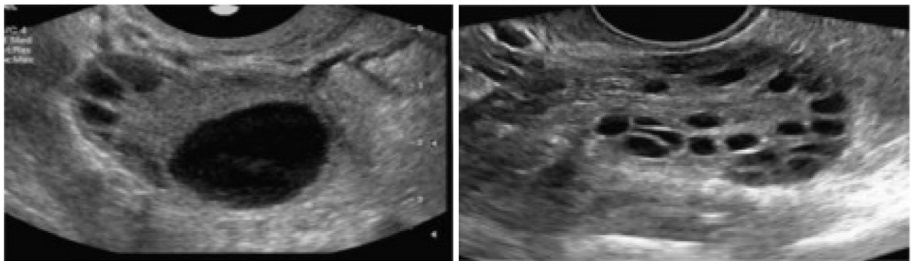


Fig. 2 Ultrasound images of normal and PCOS affected ovary

dataset must be supplemented. To work with the current dataset, users simply need to make little adjustments like flips, translations, and rotations. We refer to this procedure as data augmentation.

These would appear to the neural network to be different images. The utility of the data can be increased through data-augmentation. Data augmentation is this method of increasing the quantity and variety of data. It is not necessary to acquire new data; existing data can be changed instead. By using augmentation, it is possible to stop the neural network from picking up useless patterns, effectively improving performance. The image data can be altered and improved in a variety of ways. The most popular procedures include: 1. Rotation, which rotates the image in a random direction. 2. Shifting: To move a picture, the translation range must be manually provided, and the image is shifted to the left or right. 3. Cropping entails reducing the original image's size and resizing it to a particular resolution. 4. Flipping: The image is rotated 180 degrees horizontally. 5. Altering the brightness level: Images are alternately made brighter and darker at random.

Data augmentation plays a crucial role in training Convolutional Neural Networks (CNNs) for medical image analysis by artificially increasing the diversity of the training dataset through various transformations such as rotations, translations, and flips. This technique enhances the model's ability to generalize and improve its performance by exposing it to a wider range of variations in the images, thus reducing the risk of overfitting. Additionally, data augmentation addresses privacy concerns by minimizing the need to share original, sensitive medical images. By generating new variations of existing images rather than collecting and distributing additional real-world data, augmentation helps protect patient confidentiality while still enabling the development of robust and effective machine learning models.

Methods for image augmentation in ultrasound images

Rotation This method involves rotating the ultrasound images by random angles. Rotation helps the model become invariant to the orientation of the images, making it more robust in recognizing features regardless of how the images are presented.

Shifting Shifting involves moving the image horizontally or vertically within a specified range. This adjustment simulates different placements of the ultrasound probe and helps the model learn to recognize features regardless of their position in the image.

Cropping Cropping reduces the size of the original image to focus on specific areas and then resizes it to the desired resolution. This technique helps the model focus on relevant parts of the image while discarding unnecessary information.

Flipping Flipping involves mirroring the image horizontally. This technique helps in creating a mirrored version of the image which is useful for learning features that are not orientation-specific.

Proposed machine learning techniques

In this research work, machine learning techniques such as LR, NB and SVM are used to diagnose PCOS using clinical data features. The results are analysed, and performance of the algorithms are evaluated based on measures like accuracy, precision, recall, F1-measure.

Machine learning models play a pivotal role in the diagnosis of Polycystic Ovary Syndrome (PCOS) by leveraging large amounts of clinical and diagnostic data to identify patterns that may not be easily recognizable through traditional methods. These models enable healthcare practitioners to analyze complex and multidimensional data, such as clinical features and ultrasound images, with high precision. By using supervised learning techniques, machine learning models can be trained on labeled datasets, where the input features (such as hormone levels, BMI, and menstrual history) are associated with known PCOS outcomes. The models then learn the underlying relationships between these features and the presence of PCOS, allowing them to make predictions about new, unseen cases.

For instance, Logistic Regression (LR) provides a straightforward approach to modeling the relationship between clinical variables and the likelihood of PCOS. However, models like Support Vector Machines (SVM) and Naïve Bayes (NB) can handle more complex interactions between features. SVM, in particular, excels at classifying cases by creating decision boundaries that separate PCOS-affected patients from non-affected ones based on their clinical characteristics. The use of machine learning allows for the automation of PCOS diagnosis, reducing the time and effort required for manual data analysis. Additionally, machine learning algorithms can improve diagnostic accuracy by uncovering subtle patterns and correlations in the data that might be overlooked by human clinicians. These models can also assist in predicting the likelihood of a patient developing PCOS, which is valuable for early intervention and personalized treatment planning. By integrating machine learning models into the diagnostic process, healthcare systems can enhance their capacity to diagnose PCOS.

with greater accuracy, speed, and scalability, making them essential tools in modern clinical decision-making.

Logistic regression

Logistic Regression is a statistical method utilized for predicting outcomes (i.e., outcomes with two possible values, such as success or failure, true or false, etc.). It is a widely used technique in traditional learning that works very well for classification problems. Equation (1) describes the logistic function, also known as the sigmoid function, that is used to model the relationship between the independent variables (also known as predictors or features) and the dependent variable (also known as the response or outcome) in logistic regression.

$$f(z) = \frac{1}{1+e^{(-z)}} \quad (1)$$

where z is the model's coefficients, or parameters, and the independent variables combined linearly. The probability that the outcome will occur is represented by the value that the logistic function returns, which ranges from 0 to 1. Finding the model coefficients that maximise the likelihood of the observed data is the aim of logistic regression. The likelihood of the observed data can be determined by the likelihood function, gives the coefficients of the model. In logistic regression, the optimization problem is made simpler by using the log-likelihood function. Equation (2) provides the definition of the log-likelihood function.

$$L(\text{beta}) = \text{SUM}(y_i * \log(p_i) + (1 - y_i) * \log(1 - p_i)) \quad (2)$$

Where y_i is the observed outcome, p_i is the predicted probability of the outcome being true, and beta are the coefficients of the model. The goal of optimization methods like gradient descent and Newton–Raphson method is to maximize the log-likelihood function. Once the coefficients of the model are estimated, it can be used for making predictions by plugging in the independent variable values and the coefficients into the logistic function. The logistic function's output can be understood as the likelihood that the result being true. A threshold value is usually set, such as 0.5, to determine the predicted class.

The log-likelihood function in logistic regression is a critical component used to estimate the parameters of the model by evaluating how well the model's predictions match the observed data. It represents the logarithm of the likelihood function, which measures the probability of observing the given data under a particular set of model parameters. Specifically, the log-likelihood function sums the logarithms of the predicted probabilities for each observed outcome, weighted by the actual outcomes. The primary role of the log-likelihood

function in model optimization is to provide a measure that can be maximized to find the optimal parameters. By maximizing the log-likelihood, the logistic regression model adjusts its coefficients to best fit the data, thereby improving the accuracy of its predictions. This optimization process often employs techniques such as gradient descent to iteratively adjust the parameters and converge on the values that maximize the log-likelihood function.

The primary goal of Logistic Regression in predicting PCOS is to estimate the probability of PCOS occurrence based on various clinical features. The logistic function plays a critical role by converting the linear combination of these features into a probability score. This probability is then used to make a classification decision, facilitating effective prediction and diagnosis of PCOS.

Naive Bayes

The Bayes theorem, which asserts that the likelihood of an event given certain evidence is equal to the prior probability of the event, is the foundation of the probabilistic classifier known as naive bayes. A Naive Bayes classifier uses a set of features as the evidence, and the class label serves as the event. Given the class name, the characteristics are assumed to be conditionally independent by a Naive Bayes classifier. This implies that a feature's existence or absence is independent of the existence or absence of any other feature.

Equation (3) can be used to determine the conditional probability of a class label as y given a feature-vector as x , depending on a collection of class labels and a set of feature vectors.

$$P(y|x) = P(x|y) * P(y)/P(x) \quad (3)$$

where $P(x|y)$ is the likelihood of the feature vector given the class label, $P(x)$ is the feature vector's marginal likelihood, and $P(y)$ is the prior probability of the class label. The class label that has the largest posterior probability is chosen to categorize a new feature vector. The posterior probability is calculated using the Eq. (4).

$$P(y|x) = P(x_1, x_2, x_3 \dots x_n|y) * P(y)/P(x_1, x_2, x_3 \dots x_n) \quad (4)$$

It is important to notice that the denominator in the equation is the same for all classes and can be ignored when comparing the different class labels.

The Naive Bayes classifier determines the class label for a given feature vector by applying Bayes' theorem, which calculates the posterior probability of each class based on the prior probability and the likelihood of the features given each class. Specifically, it computes the probability of each class label given the observed features by multiplying the prior probability of the class by the product of the conditional probabilities of each feature, assuming that

these features are conditionally independent given the class label. This assumption of conditional independence simplifies the calculation, as it allows the classifier to separately evaluate the contribution of each feature to the likelihood of the class without considering interactions between features. The class with the highest posterior probability is then selected as the predicted label for the feature vector, making the Naive Bayes classifier both efficient and effective for various classification tasks.

Support vector machine

Regression and classification issues can be employed by the supervised learning method known as Support Vector Machine (SVM). By employing an advanced technique, non-linearly separable data is converted to a higher dimensional space and made linearly separable. Finding the hyper-plane that optimally divides the different classes of SVM. One way to maximize the margin is to choose the hyperplane in a way that the distance is maximised between it and nearest datapoints from each of the class, or support vectors. The equation of the hyperplane is given by Eq. (5).

$$w^T * x + b = 0 \quad (5)$$

where the hyperplane's normal vector is denoted by w , feature vector by x and the bias term as b . The hyperplane dividing the two classes in a classification issue called the decision boundary. There are points that belong to one class on one side of the decision boundary and points to the other class on the other. For data that cannot be linearly separated, SVM employs the kernel method to move the information into a higher dimensional space. In a higher-dimensional space, the kernel technique enables us to calculate the dot product of two vectors without the need for us to first find the coordinates of the vectors in that space. Often used kernels are the polynomial, radial basis function (RBF), and linear ones. The optimization problem that needs to be solved to find the hyperplane is given by Eqs. (6) and (7).

$$\text{Minimize } \frac{1}{2} * w^T * w \quad (6)$$

subject to

$$y_i(w^T * x_i + b) \geq 1 \text{ for } i = 1, 2, \dots, n \quad (7)$$

where the bias term be b , feature vectors as x_i , class labels as y_i , and the normal vector (w) to the hyperplane. New data points can be classified by calculating the value of the decision function provided by Eq. (8) once the hyperplane has been located.

$$f(x) = w^T * x + b \quad (8)$$

The new data point is assigned to class 1 if the value of $f(x)$ is positive, and class 2 if the value is negative.

The primary objective of an SVM in classification tasks is to find the optimal hyperplane that maximizes the margin between classes, ensuring the best separation in the feature space. To handle non-linearly separable data, SVM uses kernel functions to map data into higher-dimensional spaces, where linear separation is more feasible. This approach enables SVM to effectively classify complex datasets by leveraging the power of higher-dimensional transformations and kernel functions.

The kernel trick in Support Vector Machines (SVM) is significant for handling non-linear data by transforming it into a higher-dimensional space where it becomes linearly separable. This technique allows SVM to apply linear classification methods to data that is not linearly separable in its original space. The kernel trick works by computing the dot product of data points in this higher-dimensional space without explicitly performing the transformation, thus saving computational resources. Commonly used kernels, such as the polynomial and radial basis function (RBF) kernels, implicitly map the original data into a new feature space where a linear hyperplane can effectively separate the different classes. By utilizing this approach, SVMs can efficiently classify complex data patterns that would otherwise be difficult to separate using traditional linear methods.

Proposed CNN based deep learning technique

A particular kind of neural network that excels at processing images is the convolutional neural network (CNN). The distinctive feature of a CNN is the use of convolutional layers, which are intended to automatically and adaptively accumulate spatial hierarchies of information from input images. The convolutional layer forms the basis of a CNN, which processes the input image through several filters. Every filter is a tiny matrix that is "convolved," or moved, in a sliding window pattern across the image to compute the dot products between each filter entry and the input image at each place. This produces a set of maps with features that are fed into the network's subsequent layer. After the convolutional layers, the feature maps are down sampled by the pooling layers, which take the maximum or average value of small, non-overlapping regions. By doing this, the feature maps' dimensionality is decreased, increasing the network's computational efficiency and lowering overfitting. The feature maps are then sent to One or more layers that are fully connected, which carry out the final classification, after one or more convolutional and pooling layers.

CNNs are commonly used in image classification tasks, such as recognizing objects in an image, facial recognition, and medical image analysis. They have also been

used for tasks like image generation, object detection, and video analysis. CNNs are widely used in image classification tasks such as recognizing objects in an image, facial recognition, and medical image analysis. The current research aims to compare the effectiveness of well-known deep learning approaches. The model is first trained using CNN, and the significance of drop out processes is then demonstrated by using the regularization method. The CNN model is then used with a different technique called data augmentation. The dataset will be obtained and split into train, validation, and test sets to conduct this study. After that data augmentation will only be used on the train set. The PCOS ultrasound photos in the collection are extremely private and have privacy concerns. Using measuring factors such as area, perimeter, extent, solidity, and orientation, data augmentation and picture processing will be carried out. Speckle noise will be removed from photos during preprocessing. The cyst section, which is the region of interest, will next be segmented using a watershed method, which accurately segments the cyst. Once the cyst has been segmented, which is necessary for the classification phase, features will be extracted from the cyst. The retrieved features and training ultrasound images will be used to train the CNN model.

Convolutional layers in CNNs are responsible for detecting and learning features from images by applying filters that extract local patterns and hierarchies. Pooling layers complement this by reducing the spatial dimensions of the feature maps, enhancing computational efficiency, preventing overfitting, and providing a more abstract representation of the features. Together, these layers enable CNNs to efficiently and effectively perform tasks such as image classification, object detection, and medical image analysis.

Proposed VGG16 pretrained model

A model that has been constructed and trained for one job is used again on another task to enhance the performance of the model in optimization. This method is known as transfer learning and is based on the pre-training technique. The trained models can be applied in one of two methods when a new model is created, or an old model is reused to extract features making use of a trained model or to finetune the model that has already been trained. A model known as VGG-16 (Oxford's Visual Geometry Group) that has already been developed by applying the ImageNet base dataset is used in the suggested method. The learned model will be applied to a brand-new dataset made up of ultrasound scans of people with and without PCOS. The VGG-16 model, which comprises of 16 convolutional and fully connected layers,

was built using the ImageNet database, which was created for picture recognition and classification.

Convolutional neural networks were designed by Oxford University's Visual Geometry Group (VGG16). This well-known CNN architecture has achieved state-of-the-art performance on multiple benchmark datasets and is frequently used for image classification applications. Using the transfer-learning method, the trained model can serve as the foundation for a new assignment. Regarding VGG16, the model has already undergone training using a sizable dataset of images and has amassed a wealth of features that may be applied as a foundation for future tasks. When utilizing VGG16, there are two primary methods for transfer learning. The trained VGG16 model is employed as a feature extractor in the first method, which is the feature extraction strategy. Convolutional layer output is fed into a new classifier, and the model's fully connected layers are removed. The fresh dataset is then used to train this new classifier. When the new dataset is modest and the pre-trained model's features are broad enough to be applicable to the new task, this strategy can be helpful. In the second method, the weights of the already trained VGG16 model are further trained on the new set of data after it has been initially trained. When the fresh dataset is bigger and it is anticipated that the model would pick up task-specific features that weren't included in the pre-trained model, this approach is helpful. It is usual practice to freeze the weights of the model's early layers, which had learnt generic features, and only train the weights of the model's later layers, which had learned more task-specific information, while fine-tuning VGG16. By doing this, the risk of overfitting is reduced, and the training process is sped up.

After all the models have been applied, their accuracy is evaluated and compared. The model is evaluated using a test dataset that predicts if an image has PCOS or not. The findings indicate that the worried woman should consult a gynaecologist and begin taking medication right away. In this research, training uses 80% of the data, validation uses 10%, and testing uses the remaining 10%.

Freezing the weights of the early layers during fine-tuning of VGG16 is crucial for maintaining the generic feature extraction capabilities of the model. This approach allows the model to focus on learning task-specific features in the later layers, reduces the risk of overfitting, accelerates the training process, and preserves the valuable knowledge gained from pre-training on a large dataset. This strategy optimizes the model's performance on new tasks while leveraging the robust feature extraction capabilities of the pre-trained network.

The proposed system leverages deep learning techniques to analyze ultrasound images for PCOS

prediction by employing Convolutional Neural Networks (CNNs), specifically the VGG16 architecture. Initially, VGG16, pre-trained on a comprehensive image dataset, is utilized to extract generic features from the ultrasound images, which include fundamental patterns and structures relevant to the task. By fine-tuning this model on a dataset of PCOS-related ultrasound images, the system adapts the pre-trained features to detect task-specific abnormalities such as cysts. This process involves freezing the weights of the early layers to retain their ability to identify basic features while updating the later layers to recognize more complex, disease-specific patterns. Additionally, data augmentation techniques, such as rotations and translations, are applied to enhance the training dataset and improve the model's robustness. This approach enables the system to accurately classify and predict PCOS based on the nuanced details present in ultrasound images, thus providing a powerful tool for early diagnosis and management of the condition.

Results and discussion

Initially a PCOS Clinical dataset of 541 instances and 45 attributes has been taken and the important features will be extracted by utilizing Pearson correlation-based feature extraction. The heatmap representation of the correlation between the features is depicted in Fig. 3. The importance of each feature is identified according to the correlation. After this process, the PCOS clinical dataset contains 541 records of 15 significant features.

The Rotterdam criteria play a crucial role in the diagnosis of Polycystic Ovary Syndrome (PCOS) by providing a standardized framework that clinicians use to identify the condition. Established in 2003 during a consensus workshop, the Rotterdam criteria define PCOS based on the presence of at least two out of three key symptoms: irregular or absent ovulation, clinical or biochemical signs of elevated androgen levels (hyperandrogenism), and polycystic ovaries visible on an ultrasound. These criteria allow for a broader range of symptom presentations, making it easier to diagnose PCOS even when all symptoms are not present simultaneously.

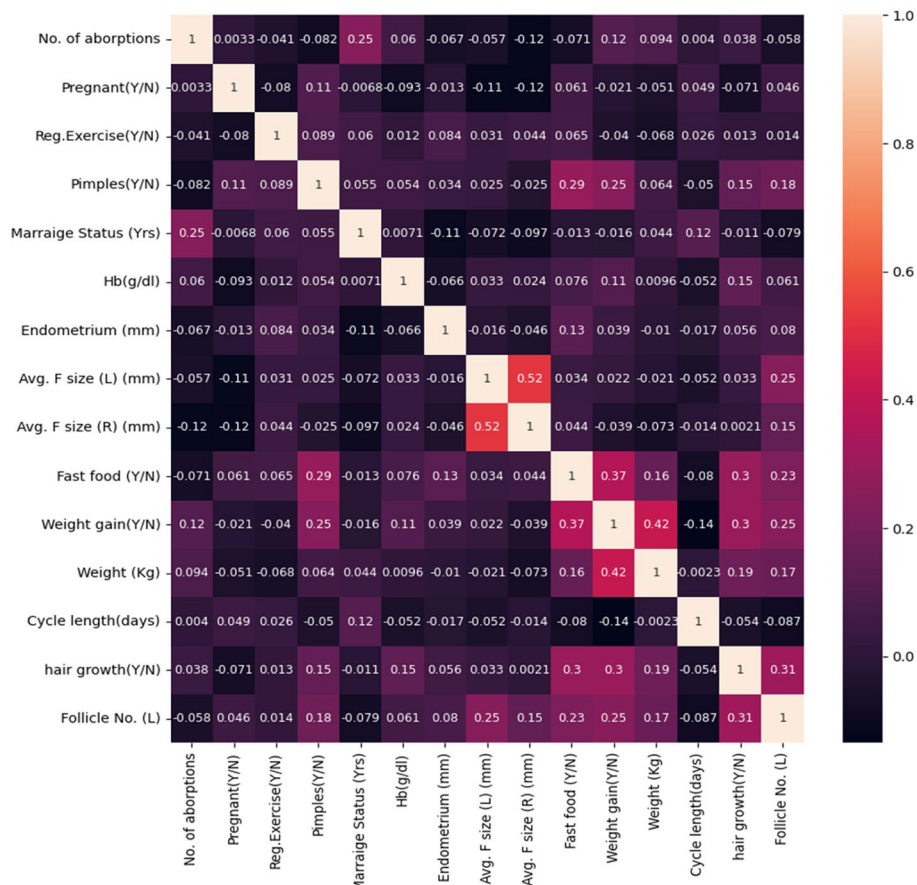


Fig. 3 Heatmap representation of highly correlated features

The justification for using the Rotterdam criteria lies in their ability to capture the heterogeneous nature of PCOS. Since the syndrome can manifest differently in each woman—some showing more pronounced reproductive symptoms, while others exhibit metabolic complications—the criteria provide flexibility. By recognizing that not all women with PCOS present with the same symptoms, the Rotterdam criteria allow for a more inclusive diagnosis. This is particularly important as it enables earlier identification of the condition in women who may not yet exhibit all the typical symptoms, facilitating earlier interventions and better management of associated risks such as infertility, diabetes, and cardiovascular disease.

Moreover, the criteria are widely accepted in clinical practice and research, ensuring consistency in diagnosis and enabling better comparisons across studies. However, some criticisms exist, particularly regarding the potential for over-diagnosis in women who have polycystic ovaries but no other significant symptoms. Despite this, the Rotterdam criteria remain a cornerstone in the diagnostic process, balancing the need for sensitivity and specificity in identifying PCOS.

To evaluate the effectiveness of a binary classification model, three often used metrics are precision, recall, and the F1-score. The model's accuracy in classifying positive and negative instances is assessed using these measures. A model's accuracy is determined by dividing its actual positively predicted value by its total positive predictions. It is calculated as in Eq. (9) where True Positives represents the positive class labels accurately identified by the model and False Positives denotes the positive class labels that were inaccurately identified by the model.

$$\text{Precision} = \text{True Positives} / (\text{True Positives} + \text{False Positives}) \quad (9)$$

A high precision score indicates that the model can accurately identify positive instances and has a low number of false positives. The percentage of actual positive instances among all of the model's accurate positive predictions is known as recall. It is calculated as in Eq. (10) where False Negatives denotes the negative class labels that were inaccurately identified by the model.

$$\text{Recall} = \text{True Positives} / (\text{True Positives} + \text{False Negatives}) \quad (10)$$

When a model has a high recall score, it means that it has minimal false negatives and can identify most positive occurrences. The F1-score is the mean of the recall and precision harmonics. It is calculated as in Eq. (11).

$$F1 - \text{Score} = 2 * (\text{Precision} * \text{Recall}) / (\text{Precision} + \text{Recall}) \quad (11)$$

Precision and recall are given equal weight in the final score by the F1-score, which is a balance between both. In some cases, it is more significant to have a high precision

score than a high recall score, and vice versa. A balance between the two is made possible by the F1-score. Performance evaluation of all the mentioned machine learning models is shown in Table 3 and this table shows the performance comparison of three different machine-learning techniques such as LR, NB and SVM on a binary classification task before applying feature extraction. Various metrics, including accuracy, precision, sensitivity, specificity, and misclassification rate, are utilized in evaluating the models' performance.

The precision of a model refers to its ability to accurately detect positive instances. SVM has the greatest precision value of 0.914286, LR has 0.885714 and NB has the lowest precision of 0.823529. This means that SVM can accurately detect positive instances. SVM has the greatest sensitivity value of 0.914286, LR has 0.861111 and NB has 0.8. This means that SVM can accurately detect positive instances. Specificity, the measure of how well the model can detect negative instances. SVM has the greatest specificity value of 0.958904, LR has 0.944444 and NB has 0.917808. This means that SVM can detect most of the negative instances. In summary, the Table 3 displays the performances comparison of different machine learning models. SVM has the highest scores in most of the metrics, namely accuracy, precision, sensitivity, and specificity. It indicates that SVM is a better performer among the three models in this specific case. The NB model is seen to be underperforming compared to other two models.

The Fig. 4 compares the Precision-Recall trade-off between LR, NB and SVM machine learning models. The result shows that best trade-off achieved by SVM model compared to LR and NB models. Three different machine-learning models are compared in Table 4 for performance comparison, after feature extraction. The models' effectiveness is assessed using several metrics such F1-score, support, accuracy, precision, recall, macro average, and weighted average. In each model 0 denotes having PCOS infected and 1 denotes PCOS not infected class labels.

Table 3 Performance evaluation of machine learning models before feature extraction

Parameters	Logistic Regression	Naïve Bayes	Support Vector Machine
Accuracy	91.67	87.96	94.44
Misclassification Rate	8.33	12.03	5.56
Precision	88.57	82.35	91.43
Sensitivity	86.11	80	91.43
Specificity	94.44	91.78	95.89

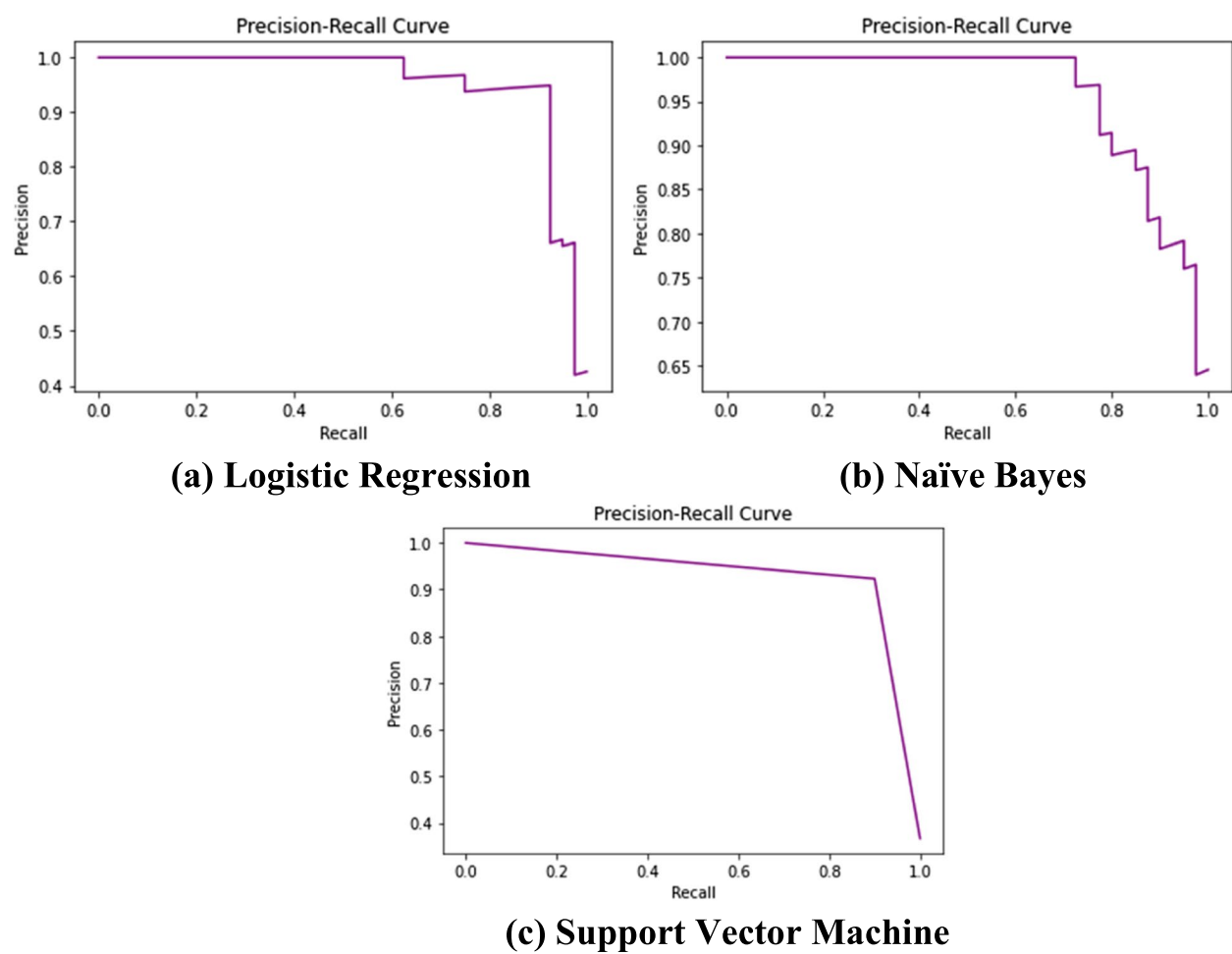


Fig. 4 Precision-recall comparison of machine learning models

Table 4 Performance evaluation of machine learning models after feature extraction

Models	LR (%)		NB (%)		SVM (%)	
	0	1	0	1	0	1
Precision	94	90	97	76	97	76
Recall	94	90	82	85	82	95
F1-Score	94	90	89	95	89	85
Support	68	41	68	41	68	41
Accuracy	93		87		87	
Macro Average	92		87		87	
Weighted Average	93		89		89	

In terms of precision, SVM has the highest scores in both classes such as 0.97 and 0.76 followed by LR with 0.94 and 0.90 and NB has the lowest scores 0.90 and 0.76 for both classes. This means that SVM can identify most of the positive instances with a highest accuracy. A model's recall refers to its ability to detect every positive case.

SVM has the highest recall value of 0.82 and 0.95 for both classes, followed by LR with 0.94 and 0.90 and NB has the lowest scores 0.90 and 0.85 for both classes. This means that SVM can detect most of the positive instances, but not as well as in precision.

SVM has the highest F1-Score of 0.89 and 0.85 for both classes, followed by LR with 0.94 and 0.90 and NB has the lowest scores 0.90 and 0.95 for both classes. This means that SVM can balance well between precision and recall, but not as well as LR and NB. Support is the number of observations for each class. It is observed that the same number of observations for classes in the case of all three models LR, NB, and SVM. The percentage of accurately identified cases is known as accuracy. Among the three models, LR having the greatest score of 0.93, SVM and NB has 0.87. This indicates that both SVM and NB models can correctly classify instances with a similar degree of accuracy. Macro Average and Weighted Average are used to calculate the average performance of model across all classes. Macro Average calculates the un-weighted mean of the metrics for each class, while Weighted Average takes the quantity of observations for each class. The table shows that all three models have similar Macro and Weighted Average scores, with LR having the highest scores of 0.92 and 0.93 respectively, followed by SVM and NB with 0.87 and 0.89 respectively.

A confusion matrix is an effective tool for assessing how well a classification algorithm performs since it gives a precise and in-depth picture of the algorithm's performance. It enables us to spot the algorithm's flaws and can be utilised to tweak the algorithm to enhance performance. It is employed to characterize the effectiveness of a categorization technique. It is used to specify how many predictions the algorithm made were true positive (TP), true negative (TN), false positive (FP), and false negative (FN). The matrix's columns reflect the projected class, while the rows represent the actual class. The number of times the algorithm accurately predicted the positive class is known as the true positive. The number of times the algorithm properly identified the negative class is known as a true negative. The number of times the algorithm mistakenly predicted the positive class is known as a false positive, and the number of times it incorrectly predicted the negative class is known as a false negative. A confusion matrix can be used to calculate several measures that includes accuracy, precision, recall, and F1 score.

A matrix of confusion in Fig. 5 shows the results of a logistic-regression algorithm that was utilized to forecast whether PCOS(polycystic ovary syndrome) would occur or not in a group of individuals. The matrix is divided into four sections, each representing a different combination of predicted and actual outcomes. In True negatives (top left): 31 individuals were correctly predicted to not have PCOS. In False positives (top right): 5 individuals were incorrectly predicted to have PCOS, when they did not. In False negatives (bottom left): 4 individuals were

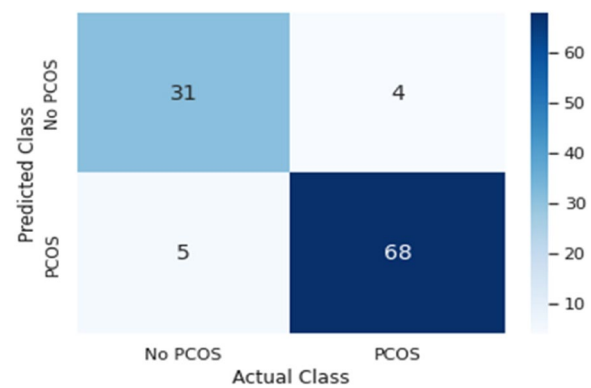


Fig. 5 Confusion matrix of logistic regression

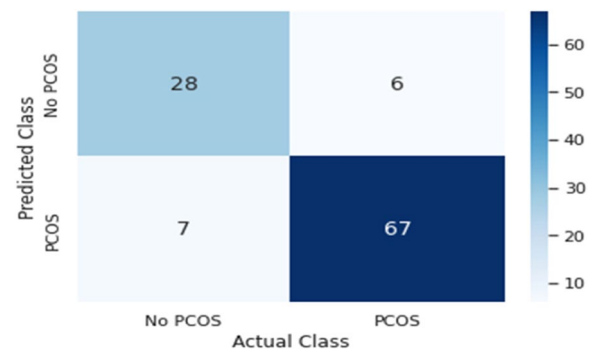


Fig. 6 Confusion matrix of Naïve Bayes

incorrectly predicted to not have PCOS, when they did. In True positives (bottom right): 68 individuals were correctly predicted to have PCOS.

The confusion matrix in Fig. 6 shows the results of a Naive Bayes model that predicts whether PCOS(polycystic ovary syndrome) would occur or not in a group of individuals. The matrix is divided into four sections, each representing a different combination of predicted and actual outcomes. In True negatives (top left): 28 individuals were correctly predicted to not have PCOS. In False positives (top right): 6 individuals were incorrectly predicted to have PCOS, when they did not. In False negatives (bottom left): 7 individuals were incorrectly predicted to not have PCOS, when they did. In True positives (bottom right): 67 individuals were correctly predicted to have PCOS. A matrix of confusion can be used as a tool for understanding the performance of a Naive Bayes model and determining areas, the model might require improvement. In this case, the model has a lower overall accuracy and specificity than the logistic regression model, but a slightly higher sensitivity and precision.

The confusion matrix in Fig. 7 shows the results of a SVM model that was used to predict PCOS (polycystic ovary syndrome) will develop or not in a group of individuals. The matrix is divided into four sections, each representing a different combination of predicted and actual outcomes. In True negatives (top left): 32 individuals were correctly predicted to not have PCOS. In False positives (top right): 3 individuals were incorrectly predicted to have PCOS, when they did not. In False negatives (bottom left): 3 individuals were incorrectly predicted to not have PCOS, when they did. In True positives (bottom right): 70 individuals were correctly predicted to have PCOS. The classification results can be interpreted using Receiver Operating Characteristics (ROC). Histograms of scores when compared True class labels of LR, NB and SVM models are depicted in Fig. 8.

We can see from the histogram shown in Fig. 8 that the score spread causes many of the positive labels to be binned close to 1, while many of the negative labels are binned close to 0. All the bins on its left and right will be labeled as 0 s and 1 s, respectively, when we place a threshold on the score. It is also observed that there are a few outliers, such as negative samples that received high scores from our model and positive samples that received low scores as in LR and NB models. These outliers will turn into false positives and false negatives, respectively, if we choose a threshold that is exactly in the middle as in SVM model.

Figure 9 illustrates how as thresholds are changed in LR, NB and SVM models, both the number of true positives and the number of quantities of positive will alter, either increasing or decreasing. As you can see, the model appears to work rather well because, as the threshold is raised, the true positive rate declines gradually while the false positive rate suddenly decreases.

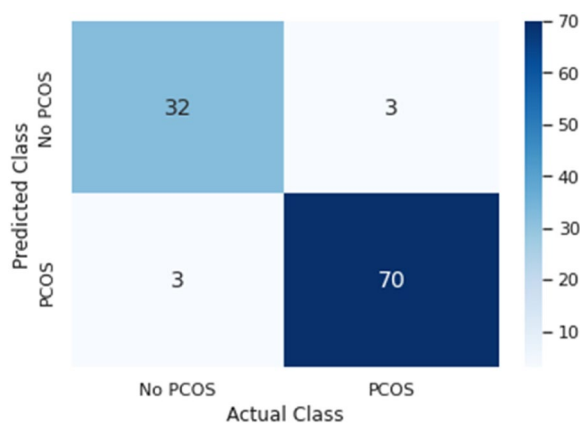


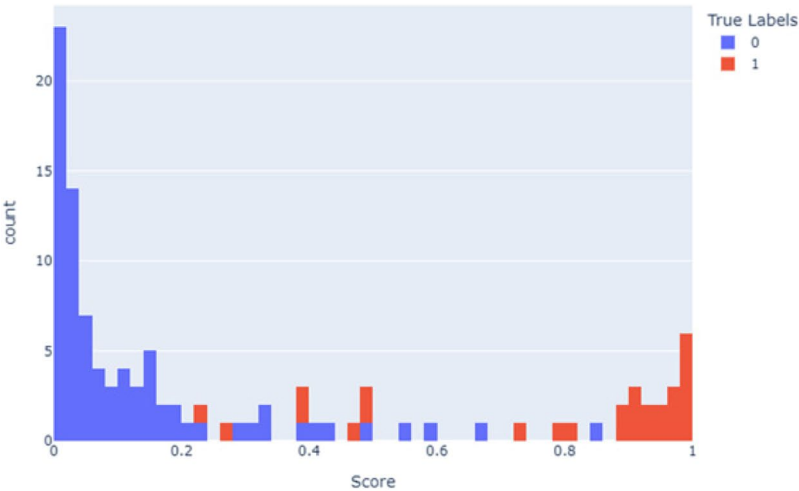
Fig. 7 Confusion matrix of support vector machine

Each of those two lines represents a different aspect of the ROC curve.

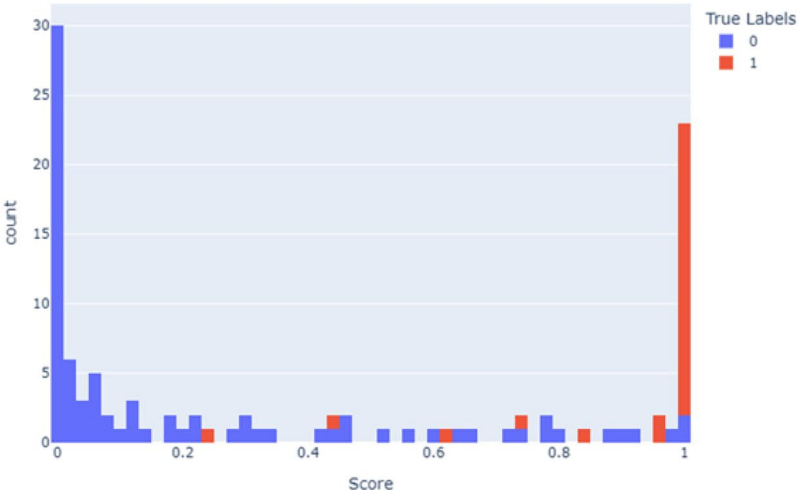
Figure 10 shows a binary ROC curve that resembles the True Positive Rate with the exception that all LR, NB, and SVM machine learning models produce distorted and flipped lines along the x-axis instead of a threshold. Table 5 describes parameters of a CNN based deep learning model and the shape of the output after each layer.

The first layer is a 2D convolutional layer (conv2d) with 10 filters and a kernel size of (3,3). The input shape is not specified, but the output shape is (None, 220, 220, 10), which means that the output is a 4-dimensional tensor with dimensions of (batch size, height, width, number of filters). The parameters size for this layer is 760. The second layer is a 2D maxpooling layer (max_pooling2d) with a pool size of (2,2). The input shape is the output of the previous layer, and the output shape is (None, 55, 55, 10) which means that the output is a 4-dimensional tensor with dimensions of (batch size, height, width, number of filters). The number of parameters for this layer is 0.

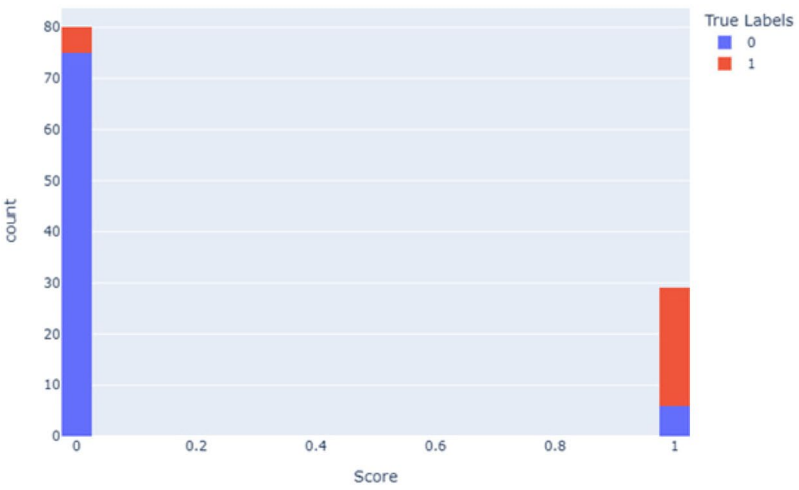
The third layer is another 2D convolutional layer (conv2d_1) with 12 filters and a kernel size of (3,3). The input shape is the output of the previous layer, and the output shape is (None, 51, 51, 12), which means that the output is a 4-dimensional tensor with dimensions of (batch size, height, width, number of filters). The number of parameters for this layer is 3012. The fourth layer is another 2D max pooling layer (max_pooling2d_1) with a pool size of (2,2). The input shape is the output of the previous layer, and the output shape is (None, 12, 12, 12) which means that the output is a 4-dimensional tensor with dimensions of (batch size, height, width, number of filters). This layer has zero parameters. The fifth layer is a flatten layer (flatten) which flattens the previous layers output into a 1-dimensional tensor. The input shape is the output of the previous layer, and the output shape is (None, 1728) which means that the output is a 1-dimensional tensor with a length of 1728. The number of parameters for this layer is 0. The sixth layer is the dense layer (dense) with 2 neurons. The input shape is the output of the previous layer, and the output shape is (None, 2) which means that the output is a 2-dimensional tensor with dimensions of (batch size, number of neurons). The parameters size for this layer is 3458. The total number of parameters for the entire model is 7,230 and all of them are trainable. There are no non-trainable parameters. The following Fig. 11 depicts the comparing training to validation accuracy and training to validation loss for different epochs of CNN based deep learning model. The result shows that at epoch 50, the lowest loss validation of 0.14 and highest accuracy validation of 0.97 are obtained by the CNN based deep-learning model.



(a) Logistic Regression

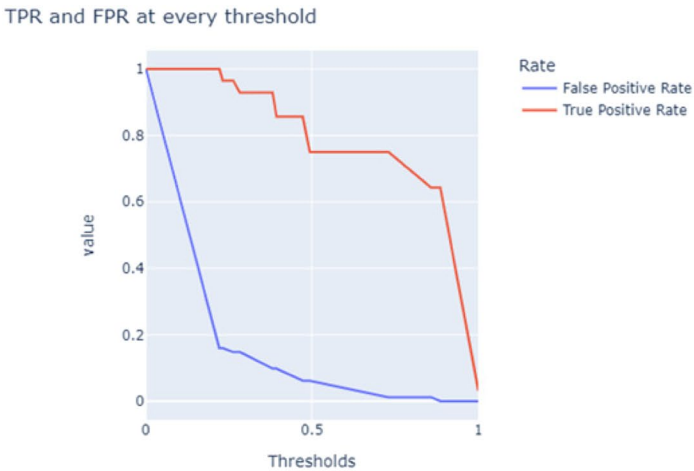


(b) Naïve Bayes

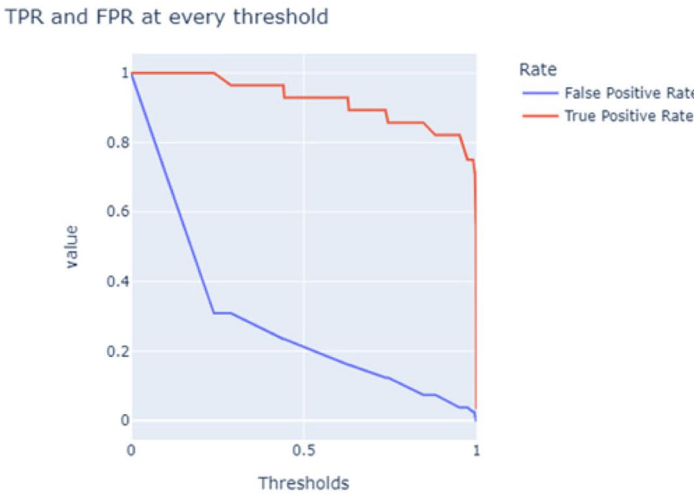


(c) Support Vector Machine

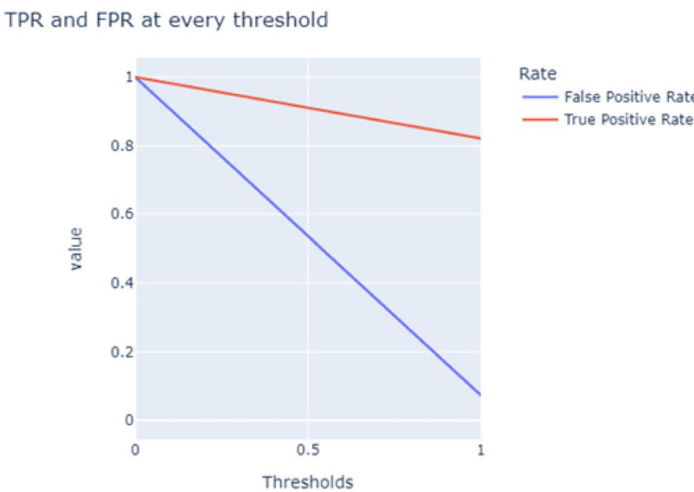
Fig. 8 Histograms of scores in machine learning models



(a) Logistic Regression



(b) Naïve Bayes



(c) Support Vector Machine

Fig. 9 Machine learning Models performance Evaluation at various Thresholds

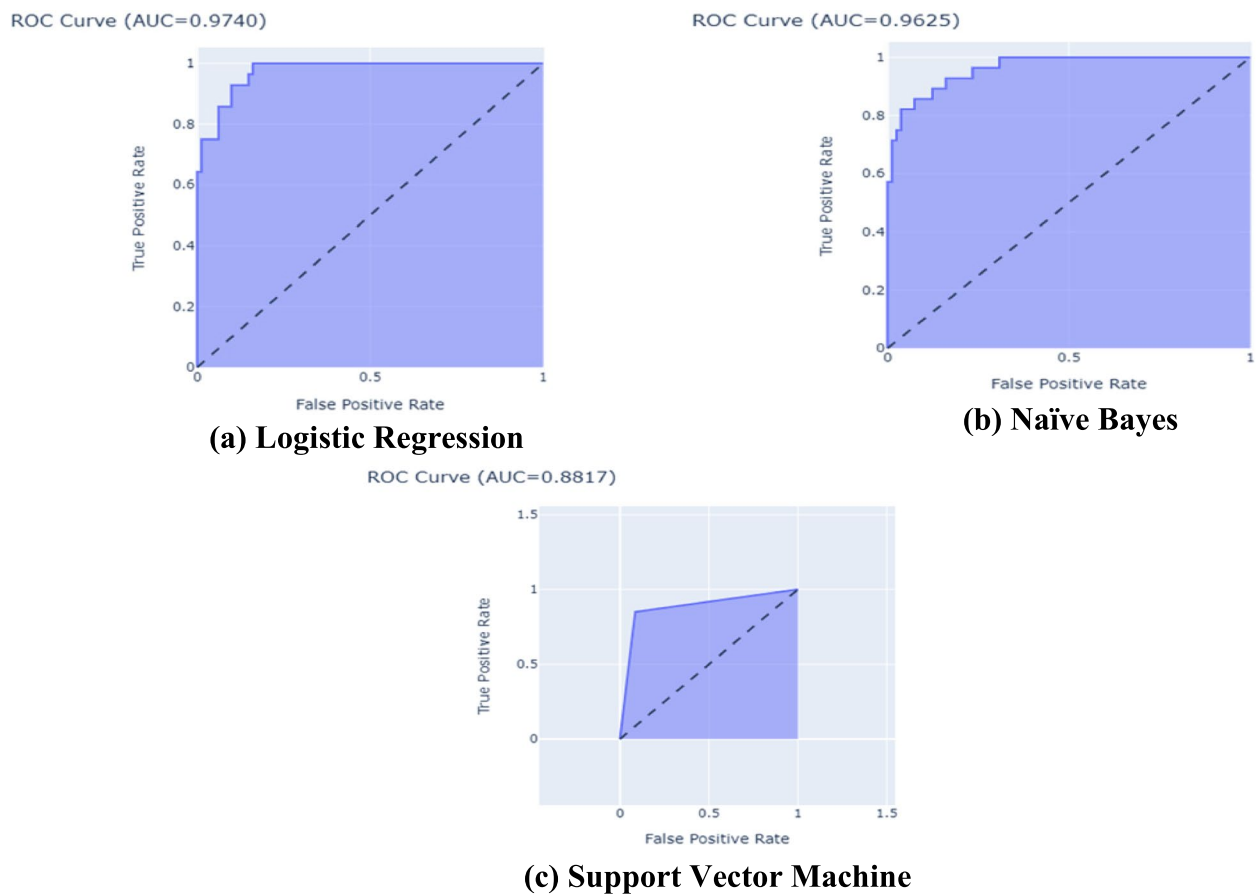


Fig. 10 ROC curve of machine learning models

Table 5 CNN based deep learning model parameters

Layer (Type)	Output Shape	Param #
conv2d (Conv2D)	(None, 220, 220, 10)	760
max_pooling2d (MaxPooling2D)	(None, 55, 55, 10)	0
conv2d_1 (Conv2D)	(None, 51, 51, 12)	3012
max_pooling2d_1 (MaxPooling2D)	(None, 12, 12, 12)	0
flatten (Flatten)	(None, 1728)	0
dense (Dense)	(None, 2)	3458
Total params: 7,230		
Trainable params: 7,230		
Non-trainable params: 0		

In addition to the machine learning and deep learning models, the ultrasound images dataset has also been trained and tested using VGG16, transfer learning-based model. For improving the classification accuracy, we have finetuned a few top layers of VGG16 that is the weights of these layers have been retrained. The results of the VGG16 transfer layer are presented in Table 6.

Table 6 shows the performance metrics of VGG16 model transfer learning model. The metrics used are Precision, Recall, F1-Score, Support, and Accuracy. In this table, the model's performance has improved with fine-tuning, with an increase in all metrics. The Macro Average and Weighted Average show a similar improvement in performance. The following Fig. 12 depicts the comparison between training and validation loss, training and validation accuracy for different epochs of VGG 16 model. The result shows that at epoch 60, the lowest validation loss of 0.08 and highest validation accuracy of 0.98 are obtained by the VGG16 transfer learning model.

The following Table 7 depicts the suggested models' performance compared to the most advanced machine learning models. The result shows that proposed deep transfer learning VGG16 model outperforms all other models for the taken dataset.

VGG16 outperforms traditional machine learning models like Logistic Regression, SVM, and even other CNN architectures due to its deep and sophisticated architecture, which includes 16 convolutional layers and fully connected layers. This depth allows VGG16

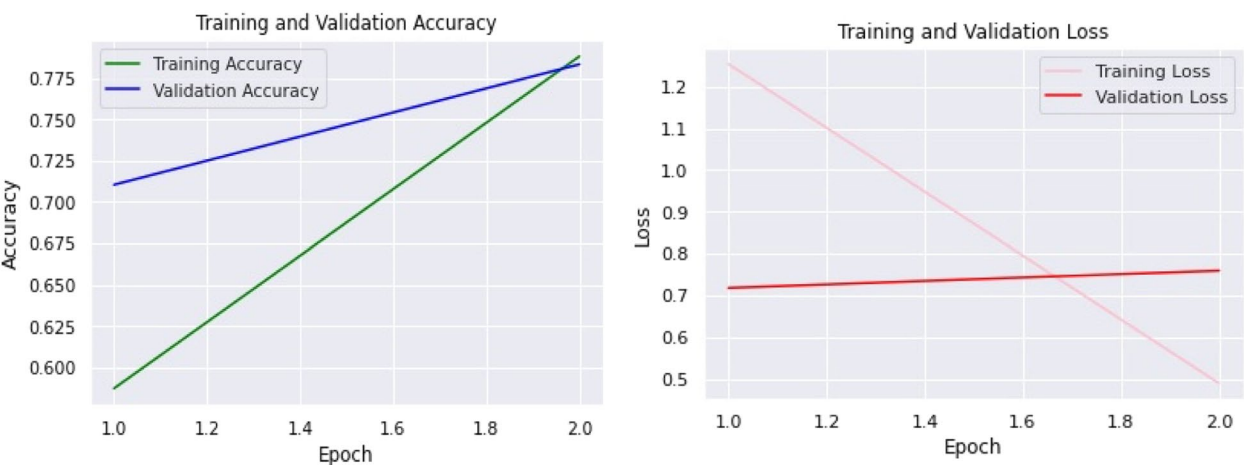


Fig. 11 CNN based deep learning model- training vs validation loss and accuracy

Table 6 Performance of VGG16 – transfer learning model

Metrics/Model	VGG16		
	0	1	
Precision	97.5	98.8	
Recall	98.3	98.2	
F1-Score	97.9	98.6	
Support	787	1145	
Accuracy	98.288		
	Precision	Recall	F1-Score
Macro Average	98.172677	98.28676	98.22867
Weighted Average	98.288382	98.28833	98.28737

to learn and capture complex hierarchical features from images, enabling it to achieve high accuracy in image classification tasks. Unlike simpler models, VGG16 benefits from extensive pre-training on large datasets, such as ImageNet, which provides a rich set of learned features that can be effectively transferred to new tasks. This pre-training enhances its ability to recognize intricate patterns and details in images, leading to superior performance compared to models with shallower architectures or those trained from scratch. Additionally, VGG16’s uniform architecture and use of small 3×3 convolutional filters contribute to its ability to

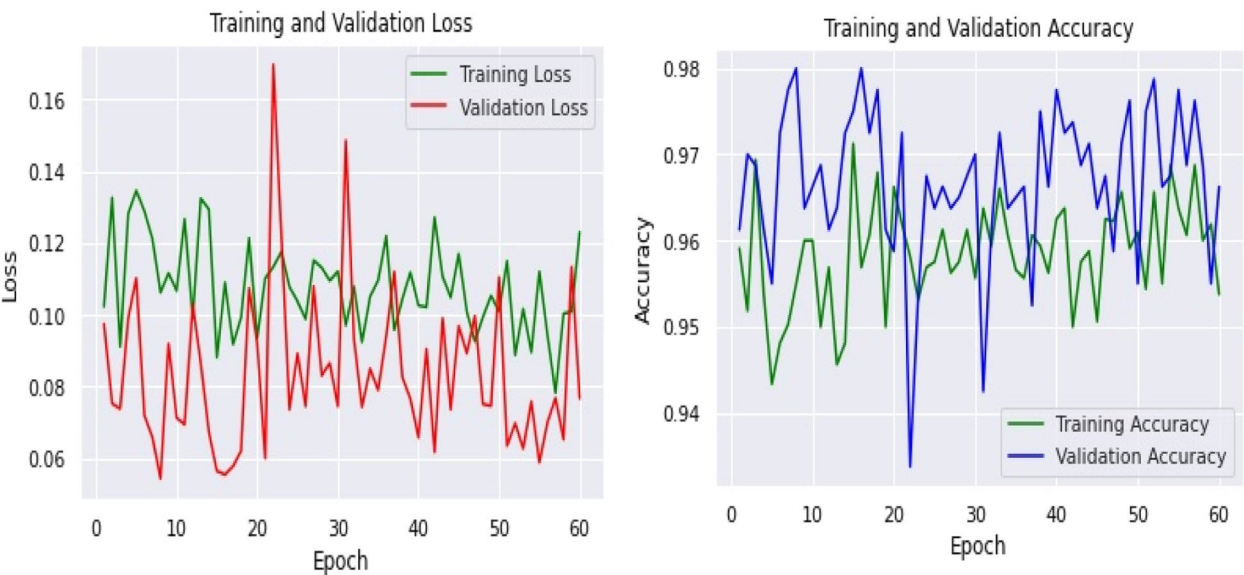


Fig. 12 Training vs validation loss and accuracy of VGG16—transfer learning model

Table 7 Performance comparison based on accuracy

Model	Accuracy
LR	93
NB	87
SVM	87
CNN	97
VGG16	98

effectively extract and utilize features from diverse and complex image datasets.

Conclusion and future work

This research work goal to conduct analysis in an efficient manner by considering PCOS dataset with clinical features and ultrasound images. Enhancing the diagnosis of PCOS requires a multifaceted approach that considers both clinical data with ultrasound imaging for a comprehensive analysis. By combining information from both modalities, a more accurate and nuanced understanding of PCOS manifestations can be attained, facilitating early detection and intervention. Clinical data provides insights into hormonal imbalances, menstrual irregularities, and other symptoms associated with PCOS, while ultrasound imaging offers visual confirmation of ovarian abnormalities such as cysts. These two sources of information enable healthcare providers to make more informed decisions regarding diagnosis and treatment, ultimately improving patient outcomes and quality of care. To achieve this goal, Feature selection based on Pearson correlation is applied over the PCOS clinical dataset and important features are identified. After that machine learning models such as LR, NB and SVM are applied to reduced features. By evaluating the performance of machine learning models, it shows that highest accuracy of 94.44% is obtained by SVM model. Next, CNN based deep learning and VGG16 transfer learning methods are applied to augmented ultrasound images dataset. The result indicates that among the proposed models, VGG16 transfer learning model produces highest accuracy of 98.288% than CNN based deep learning model. These methods support the usefulness of crucial clinical features such as follicles, hair growth and weight gain and ultrasound images for the accurate and early detection of PCOS. This will lessen the significant burden and save PCOS patients' time by lowering the number of clinical tests needed for diagnosis. In future, attention mechanism-based techniques may be incorporated in order to improve the accuracy of PCOS detection process.

Acknowledgements

This research was supported by the Researchers Supporting Project number (RSPD2024R846), King Saud University, Riyadh, Saudi Arabia.

Authors' contributions

K.S took care of the review of literature and methodology. M.D.M.S has done the formal analysis, data collection and investigation. M.T.R and T.A have done the initial drafting and statistical analysis. N.A.A and T.E.Y have supervised the overall project. All the authors of the article have read and approved the final article.

Funding

This research was supported by the Researchers Supporting Project number (RSPD2024R846), King Saud University, Riyadh, Saudi Arabia.

Availability of data and materials

The datasets generated and/or analysed during the current study are available in the Kaggle repository, [<https://www.kaggle.com/datasets/prasoonkottaratil/polycystic-ovary-syndrome-pcos>] [<https://www.kaggle.com/datasets/anaghachoudhari/pcos-detection-using-ultrasound-images>].

Declarations

Ethics approval and consent to participate

Not applicable as the research is done on the publicly available dataset.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Department of Artificial Intelligence, Kongu Engineering College, Erode, India. ²Vel Tech Rangarajan Dr.Sagunthala R&D Institute of Science and Technology, Chennai, India. ³Department of Computer Science and Engineering JAIN (Deemed-to-Be University), Bengaluru 562112, India. ⁴Computer Skills, Department of Self-Development Skill, Common First Year Deanship, King Saud University, Riyadh 11362, Saudi Arabia. ⁵Computer Science Department, College of Computer and Information Sciences, King Saud University, Riyadh 12372, Saudi Arabia. ⁶Department of Mathematics, Dilla University, Dilla, Ethiopia.

Received: 26 July 2024 Accepted: 18 September 2024

Published online: 01 October 2024

References

1. Ndefo UA, Eaton A, Green MR. Polycystic ovary syndrome: a review of treatment options with a focus on pharmacological approaches. *P.T.* 2013;38(6):336–55 PMID: 23946629; PMCID: PMC3737989.
2. Witchel SF, Oberfield SE, Peña AS. Polycystic ovary syndrome: pathophysiology, presentation, and treatment with emphasis on adolescent girls. *J Endocr Soc.* 2019;3(8):1545–73. <https://doi.org/10.1210/je.2019-00078>. PMID:31384717;PMCID:PMC6676075.
3. Lujan ME, Chizen DR, Pierson RA. Diagnostic criteria for polycystic ovary syndrome: pitfalls and controversies. *J Obstet Gynaecol Can.* 2008;30(8):671–9. [https://doi.org/10.1016/S1701-2163\(16\)32915-2](https://doi.org/10.1016/S1701-2163(16)32915-2). PMID: 18786289;PMCID:PMC2893212.
4. Salman Hosain AKM, Mehedi MHK, Kabir IE. PCONet: A Convolutional Neural Network Architecture to Detect Polycystic Ovary Syndrome (PCOS) from Ovarian Ultrasound Images. 2022 International Conference on Engineering and Emerging Technologies (ICEET). Kuala Lumpur: IEEE; 2022. <https://doi.org/10.1109/iceet56468.2022.10007353>.
5. Purnama B, Wisesti UN, Nhita F, Gayatri A, Mutiah T. A classification of polycystic Ovary Syndrome based on follicle detection of ultrasound images, In: 2015 3rd International Conference on Information and Communication Technology (IColCT). 2015. p. 396–401. <https://doi.org/10.1109/IColCT.2015.7231458>.

6. Jegadeesan S, Kamalesh S, Aishwarya SS, Harshini RG. Developing a decision tree model to diagnose polycystic ovary syndrome and evaluating it using diverse machine learning techniques. *Telematique*. 2023;22(01):1930–7.
7. Dewi RM, Adiwijaya, Wisesty UN, Jondri. Classification of Polycystic Ovary Based on Ultrasound Images Using Competitive Neural Network. *J Phys Conf Ser*. 2018;971:012005. <https://doi.org/10.1088/1742-6596/971/1/012005>.
8. Abouhawwash M. Automatic Diagnosis of Polycystic Ovarian Syndrome Using Wrapper Methodology with Deep Learning Techniques. *Comput Syst Sci Eng*. 2023;47(1):239–53. <https://doi.org/10.32604/csse.2023.037812>.
9. Alamoudi A, Khan IU, Aslam N, Alqahtani N, Alsaif HS, Al Dandan O, Al Gadeeb M, Al Bahrani R. A Deep learning fusion approach to diagnosis the Polycystic Ovary Syndrome (PCOS). *Appl Comput Intell Soft Comput*. 2023;2023:9686697.
10. Rachana B, Priyanka T, Sahana KN, Supriya TR, Parameshachari BD, Sunitha R. Detection of polycystic ovarian syndrome using follicle recognition technique. *Glob Transit Proc*. 2021;2(2):304–8.
11. Vishwakarma V, Chethan S, Datla MT, Aqib MM, Roy S, Thasni T. Prediction of Severity of Polycystic Ovarian Syndrome Using Artificial Neural Networks. *Second International Conference on Image Processing and Capsule Networks*. 2021. p. 589–98. https://doi.org/10.1007/978-3-030-84760-9_50.
12. Padmapriya B, Kesavamurthy T. Detection of follicles in polycystic ovarian syndrome in ultrasound images using morphological operations. *J Med Imaging Health Infor*. 2016;6(1):240–3.
13. Bellver J, Rodríguez-Taberner L, Robles A, Muñoz E, Martínez F, Landeras J. Polycystic ovary syndrome throughout a woman's life. *J Assist Reprod Genet*. 2018;35:25–39.
14. Wojtusik J, Michalski RS, Simanivanh T, Baranova AV. Towards application of rule learning to the meta-analysis of clinical data: an example of the metabolic syndrome. *Int J Med Inform*. 2009;78(12):e104–11.
15. Zhang N, Wang H, Xu C, Zhang L, Zang T. DeepGP: an integrated deep learning method for endocrine disease gene prediction using omics data. *Front Cell Dev Biol*. 2021;9:700061.
16. Tan J, Wang QY, Feng GM, Li XY, Huang W. Increased risk of psychiatric disorders in women with polycystic ovary syndrome in Southwest China. *Chin Med J (Engl)*. 2017;130:262–6.
17. Merkin SS, Phy JL, Sites CK, Yang D. Environmental determinants of Polycystic ovary syndrome. *Fertil Steril*. 2016;106:16–24.
18. Januszewski M, Issat T, Jakimiuk AA, Santor-Zaczynska M, Jakimiuk AJ. Metabolic and hormonal effects of a combined Myo-inositol and d-chiro-inositol therapy on patients with polycystic ovary syndrome (PCOS). *Ginekol Pol*. 2019;90:7–10.
19. Syeda Sidra SS, Tariq MH, Farrukh MJ. Evaluation of clinical manifestations, health risks, and quality of life among women with polycystic ovary syndrome. *PLoS One*. 2019;14:e0223329.
20. Mehrotra P, Chatterjee J, Chakraborty C, Ghoshdastidar B, Ghoshdastidar S. Automated Screening of Polycystic Ovary Syndrome Using Machine Learning Techniques. 2011 Annual IEEE India Conference. Hyderabad: IEEE; 2011. <https://doi.org/10.1109/indcon.2011.6139331>.
21. Vasavi G, Jyothi DS. Polycystic ovary syndrome detection using various machine learning methods-a review. *J Adv Res Dyn Control Syst*. 2017;5:234–9.
22. Suha SA, Islam MN. An extended machine learning technique for polycystic ovary syndrome detection using ovary ultrasound image. *Sci Rep*. 2022;12:17123. <https://doi.org/10.1038/s41598-022-21724-0>.
23. Gopalakrishnan C, Iyapparaja M. Detection of Polycystic Ovary Syndrome from Ultrasound Images Using SIFT Descriptors. *Bonfring Int J Softw Eng Soft Comput*. 2019;9(2):26–30. <https://doi.org/10.9756/bijsec.9017>.
24. Vikas B, Radhika Y, Vineesha K. Detection of polycystic ovarian syndrome using convolutional neural Networks K. *Int J Curr Res Rev*. 2021;13(6):156–60.
25. Chauhan P, Patil P, Rane N, Raundale P, Kanakia H. Comparative Analysis of Machine Learning Algorithms for Prediction of PCOS. 2021 International Conference on Communication Information and Computing Technology (ICCICT). Mumbai: IEEE; 2021. <https://doi.org/10.1109/iccict50803.2021.9510101>.
26. Kshetrimayum C, Sharma A, Mishra VV, Kumar S. Polycystic ovarian syndrome: environmental/occupational, lifestyle factors; an overview. *J Turk Ger Gynecol Assoc*. 2019;20(4):255–63.
27. Sun Y, Liu X, Ding Q, Yin S, Yang H. Acupuncture combined with metformin for polycystic ovary syndrome: a protocol for systematic review and meta-analysis. *Medicine (Baltimore)*. 2022;101:32234.
28. Gyliene A, Strakysyte V, Zaboriene I. Value of ultrasonography parameters in diagnosing polycystic ovary syndrome. *Open Med (Wars)*. 2022;17(1):1114–22. <https://doi.org/10.1515/med-2022-0505>. PMID:35799603;PMCID:PMC9210988.
29. Akre S, Sharma K, Chakole S, et al. Recent advances in the management of polycystic ovary syndrome: a review article. *Cureus*. 2022;14(8):e27689. <https://doi.org/10.7759/cureus.27689>.
30. Peshawar KMU. Treatment with metformin and combination of metformin and pioglitazone in polycystic ovarian syndrome. 2021. Available online: <https://ClinicalTrials.gov/show/NCT03117517>.
31. Rashid R, Mir SA, Kareem O. Polysystic ovarian syndrome-current pharmacotherapy and clinical implications. *Taiwan J Obstet Gynecol*. 2022;61(1):40–50.
32. Chakravarthy SS, Bharanidharan N, Kumar VV, Mahesh TR, Khan SB, Almusharraf A, Albalawi E. Intelligent Recognition of Multimodal Human Activities for Personal Healthcare. *IEEE Access*. 2024;12:79776–86. <https://doi.org/10.1109/access.2024.3405471>.
33. Kujanpää L, Arffman RK, Pesonen P, Korhonen E, Karjula S, Järvelin MR, Franks S, Tapanainen JS, Morin-Papunen L, Piltonen TT. Women with polycystic ovary syndrome are burdened with multimorbidity and medication use independent of body mass index at late fertile age: a population-based study. *Acta Obstet Gynecol Scand*. 2022;101(7):728–36.
34. Kaggle URL: <https://www.kaggle.com/datasets/prasoonkottarathil/polycystic-ovary-syndrome-pcos>. Accessed 5 Jan 2024.
35. Kaggle URL: <https://www.kaggle.com/datasets/anaghachoudhari/pcos-detection-using-ultrasound-images>. Accessed 5 Jan 2024.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.