PREDICTING POLYCYSTIC OVARY SYNDROME THROUGH MACHINE LEARNING TECHNIQUE USING PATIENTS' SYMPTOM DATA AND OVARY ULTRASOUND IMAGES

SAYMA ALAM SUHA (SN. 0420140003)

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of Master of Science in Computer Science and Engineering

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
MILITARY INSTITUTE OF SCIENCE AND TECHNOLOGY
DHAKA, BANGLADESH

DECEMBER 2022

PREDICTING POLYCYSTIC OVARY SYNDROME THROUGH MACHINE LEARNING TECHNIQUE USING PATIENTS' SYMPTOM DATA AND OVARY ULTRASOUND IMAGES

M.Sc. Engineering Thesis

By

SAYMA ALAM SUHA (SN. 0420140003)

Approved as to style and content by the Board of Examination on 20 September 2022:

Lt Col Muhammad Nazrul Islam, PhD Associate Professor of Computer Science and Engineering MIST, Dhaka	Chairman (Supervisor) Board of Examination
Dr. Md. Forhad Rabbi Professor of Computer Science and Engineering SUST, Sylhet	Member (External) Board of Examination
Dr. Md. Mahbubur Rahman Professor of Computer Science and Engineering MIST, Dhaka	Member Board of Examination
Brig Gen Abdur Razzak, SUP, psc Professor of Computer Science and Engineering MIST, Dhaka	Head of the Dept. Member (Ex-officio)

Department of Computer Science and Engineering, MIST, Dhaka

PREDICTING POLYCYSTIC OVARY SYNDROME THROUGH MACHINE LEARNING TECHNIQUE USING PATIENTS' SYMPTOM DATA AND OVARY ULTRASOUND IMAGES

DECLARATION

I hereby declare that the study reported in this thesis entitled as above is my own original work and has not been submitted before anywhere for any degree or other purposes. Further I certify that the intellectual content of this thesis is the product of my own work and that all the assistance received in preparing this thesis and sources have been acknowledged and/or cited in the reference Section.

Sayma Alam Suha

Department of Computer Science and Engineering, MIST, Dhaka.

ABSTRACT

Polycystic ovary syndrome (PCOS) is the most prevalent endocrinological abnormality & one of the primary causes of anovulatory infertility in women globally. The real-world clinical PCOS detection technique is critical since the accuracy of interpretations being substantially dependent on vast numbers of symptoms & physician's expertise. An artificially intelligent PCOS detection system might be a feasible alternative to the typical diagnostic procedure. Thus, the objectives of this study are: to propose intelligent computer-aided PCOS detection techniques based on patient symptom data & ovary ultrasonography(USG) images; as well as to compare the performances of the proposed techniques with the existing machine learning (ML) based methodologies. To achieve these objectives, firstly, a modified ensemble ML classification technique has been proposed for PCOS detection with patients' symptom data utilizing state-of-the-art stacking technique with five traditional ML models as base learners & one bagging or boosting ensemble model as meta-learner. At this phase, three distinct types of feature selection methods are applied to explore the minimal & optimal features for PCOS detection. Secondly, for PCOS prediction using ovary USG images, an extended ML classification technique has been proposed, trained & tested over 594 ovary USG images; where the Convolutional Neural Network (CNN) with different state-of-the-art techniques & transfer learning has been employed for feature extraction from the images; then stacking ensemble machine learning technique using conventional models as base learners & bagging or boosting ensemble model as meta-learner have been used on that reduced feature set to classify between PCOS & non-PCOS ovaries. Finally, the comparative analysis has revealed that the proposed techniques for both cases significantly enhances the predictive performances in comparison to the existing ML based techniques. In case of symptom data, the proposed ensemble technique with 'Gradient Boosting' classifier as meta learner outperforms others with 95.7% accuracy while using the features selected using PCA method. Using the proposed technique with USG images, the best performing results are obtained by incorporating the 'VGGNet16' pre-trained model with CNN architecture as feature extractor and then stacking ensemble model with the meta-learner being 'XGBoost' model as image classifier with an accuracy of 99.89% for classification.

ACKNOWLEDGEMENTS

All praise and glory to Almighty Allah, who provided me with the courage, health, and patience to complete this research work.

First and foremost I am extremely grateful to my thesis supervisor Lt Col Muhammad Nazrul Islam, PhD; Associate Professor (Instructor Class-A), Department of Computer Science & Engineering (CSE), MIST. His immense knowledge and plentiful experience have encouraged me in all the time of my academic research and daily life. He consistently allowed this research to be my own work, but steered me in the right direction whenever I needed it.

I'm extremely grateful to the Information and Communication Technology (ICT) Division under the Ministry of Posts, Telecommunications and Information Technology for enriching the study through granting fellowship. I would like to express my deepest gratitude to one and all of CSE Dept, MIST; who directly or indirectly, have lent their hand in this work. Also I am thankful to all my coursemates for their kind support.

Finally, I must express my very profound gratitude to my family members for providing me with unfailing motivation and continuous encouragement throughout my years of study and through the process of conducting my research and conducting this thesis. This endeavor would not have been possible without them.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATION

PCOS	: Polycystic Ovary Syndrome
ML	: Machine Learning
Ultrasonography	: USG
Support Vector Machine	: SVM
Decision Tree	: DT
Naive Bayes	: NB
K Nearest Neighbour	: KNN
Random Forest	: RF
Convolutional Neural Network	: CNN
Deep Learning	: DL
WHO	: World Health Organization

CHAPTER 1

INTRODUCTION

This chapter comprises of the thesis background that includes an overview of the key research topic. Here, the motivation and problem statements of the thesis have been highlighted. Following that, the thesis objectives are stated point by point. Next, the overall methodological framework of the thesis has been presented. And finally, the organization of the remaining chapters of the thesis are described.

1.1 Thesis Background

Polycystic ovary syndrome (PCOS) is one of the most widespread endocrinological anomalies affecting one out of every ten pre-menopausal reproductive women worldwide, being one of the leading causes of infertility. PCOS is associated with excessive rise in male androgen hormone in the female body, which causes a persistent disruption in hormonal levels and as a consequence, adversely affects normal ovarian functions, resulting in the growth of numerous cysts inside the ovary (Ajmal, S. Z. Khan, and Shaikh 2019). According to epidemiological research, PCOS is found to yield a number of detrimental life-threatning impacts that are prevalent in PCOS patients, with 44–70% women suffering from various critical side effects as well as affecting one in every ten premenopausal reproductive female throughout the world. PCOS has been recognized as the leading cause of anovulatory infertility as well as being related to a range of metabolic and psychological disorders; including irregular menstrual periods, hirsutism, abrupt obesity, type 2 diabetes, thyroid abnormalities, increased depression, sexual dissatisfaction, etc. lowering the quality of a healthy way of life (Palomba, Piltonen, and Giudice 2021; Kałużna et al. 2020). Multiple studies have also revealed that, women having PCOS are at higher risk of suffering from endometrial and ovarian cancer which can lead to death if not detected early (Jia et al. 2020; Meczekalski et al. 2020).

However, accumulating evidence suggests that if a well-standardized diagnostic approach can be used to identify PCOS in a timely manner, the condition can be recovered by appropriate, symptom-oriented, long-term and dynamic treatments (Escobar-Morreale 2018). The Rotterdam criteria for polycystic ovarian syndrome (PCOS) are three criteria that are commonly used to diagnose PCOS by a wide spectrum of medical practitioners; those are: hyperandrogenism, menstrual irregulations and existance of multiple cysts in ovary ultrasonography (Rui Wang and Mol 2017). Among these, the identification of numerous cysts using ultrasound scanning is the most reliable method of detecting PCOS (Balen et al. 2003).

1.2 Motivation and Problem Statement

PCOS is a condition that is strongly linked to a wide range of other physical anomalies in the patient's body and so the detection process depends on a vast numbers of the clinical and metabolic symptoms. In that case, though observing the presence of cysts in the ovaries through USG image analysis is an effective procedure, the images captured by the radiologist also consist of the the presence of significant noises that makes the diagnosis more critical. Because of these reasons, the medical analysis for detecting PCOS becomes very time-consuming and difficult, with the risk of human mistake. Also, due to the wide range of symptoms associated with PCOS and the existence of a variety of concomitant gynecological problems, PCOS becomes exceedingly challenging to diagnose by the physicians effectively at an early stage (Arentz et al. 2021; M. N. Islam, Zavin, et al. 2017). Furthermore, the effective identification of PCOS necessitates a lot of clinical test evaluations by qualified healthcare providers as well as the observation of both gynecologists and radiologists, which is sometimes unattainable in areas where expert physicians and resources are scarce. In addition, in least developed and underdeveloped countries, there is a scarcity of experienced radiologists and gynecologists who can appropriately diagnose PCOS in a timely manner. Therefore, The traditional clinical approach of PCOS detection requires sophisticated procedures with supervision of expert clinicians (Gynecologist and Radiologist) which may become difficult to manage. As a result, many young women who are suffering from this serious condition go undetected and untreated for long periods of time which worsen their conditions with enhanced risk of fatality.

Thus, researchers worldwide are now working to develop effective PCOS detection approach that would employ a variety of modern computational techniques. The automatic identification of PCOS using different computer-assisted approaches has recently gained the attention of researchers as a solution to this issue. But, a limited number of studies have been conducted for efficient prediction and diagnosis of PCOS using advanced computational techniques. Also, few studies have been performed with an aim to explore the minimal and most significant features for PCOS detection. Moreover, women's reproductive health still remains a neglected issue, with a few studies focusing on early and efficient detection of PCOS considering both patients' symptoms and image data. Most of the existing ML based studies emphasized on predicting PCOS using either symptoms data or USG images, exploring the optimum set of features, finding out the best performing classifier by implementing traditional and ensemble (bagging, boosting) ML models. As such, an integrated or extended ML based approach may enhance the prediction performance and reduce the computational complexity to predict PCOS using both types of data (symptom data and image data). Thus, further investigation is required to explore extended ML classifiers for improving the current ML based models by incorporating different state-of-the-art ML techniques like transfer learning, stacking, etc.

1.3 Thesis Objectives

The objectives of this thesis are:

- To propose a multi-level stacked ensemble classifier for PCOS prediction using patients' symptom data.

- To propose an extended ML classifier by integrating traditional ML with transfer learning based neural network for PCOS prediction using USG images.

- To compare the performances and computational complexities of the proposed classifiers with the existing ML based models for predicting PCOS.

1.4 Methodological Overview

To attain the objectives of the thesis, the methodology of the research work has been broadly divided into three phases. At initial phase, the existing relevant studies for PCOS prediction using various machine learning techniques have been thoroughly analyzed and thus a systematic literature review has been formulated in order to attain a complete knowledge in this domain. In the second phase, using the patients' clinical symptom dataset, a stacked ensemble machine learning classifier has been proposed, trained and tested; with an aim to fulfill research objective 1. Moreover, at this phase the performance of the proposed technique has been comparatively evaluated with the existing or previously applied ML based model for predicting PCOS with symptom data (fulfilling research objective 3). And finally, an extended hybrid machine learning classifier by integrating traditional ML with transfer learning based neural network has been proposed, trained and tested using ovary ultrasonography images; with an aim to fulfill research objective 2. At this stage also, a comparative performance analysis have been performed with the proposed technique and other existing techniques for predicting PCOS with USG image data. The framework of the methodological overview has been illustrated in Figure 1.1.

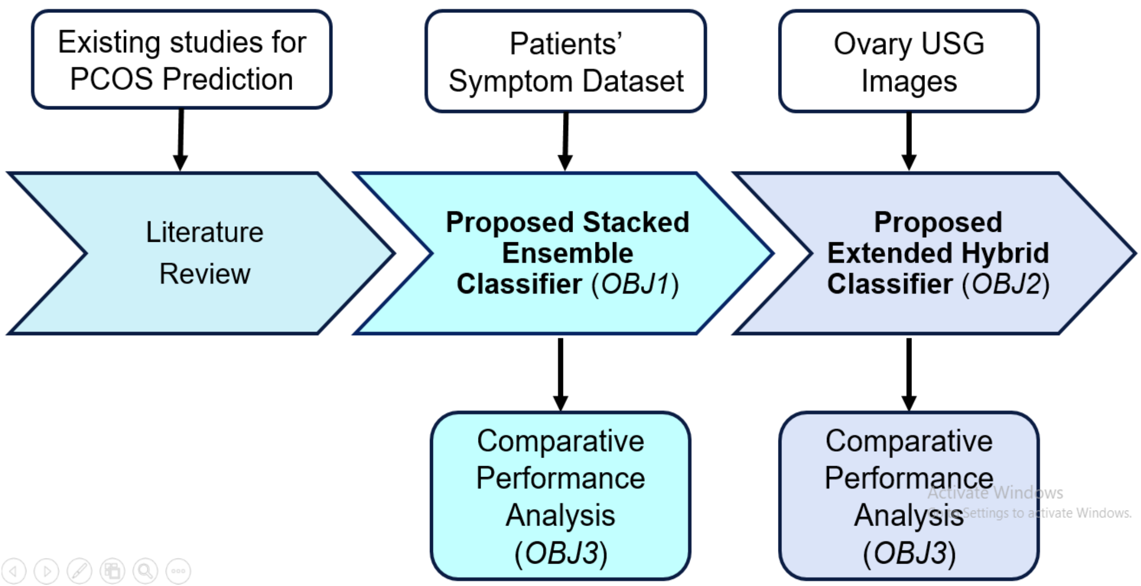


Fig. 1.1. Framework of methodological overview

1.5 Organization of the Chapters

The rest of the thesis has been structured in the following way:

Chapter 2 describes the necessary theoretical background of the study; which includes the primary discussion about the PCOS condition in female body, how automated detection of PCOS can aid in this domain, the existing works on PCOS detection using various computer-aided technologies employing both patient symptom data and ovary USG images, as well as the research gaps in these areas.

Chapter 3 includes the systematic literature review(SLR) of the related works in this domain; which demonstrate the methodology utilized for conducting the SLR, the data extraction and analysis process from the relevant studies, finally the summmarization of the review findings with potential research opportunities in this domain.

Chapter 4 depicts the methodology, result analysis and discussion of the proposed multi-level stacked ensemble ML classifier for PCOS prediction using patients' clinical symptom data as well as the comparative performance analysis of the proposed technique with other existing techniques.

Chapter 5 demonstrates the methodology, result analysis and discussion of the proposed extended ML classifier with transfer learning based neural network for PCOS prediction using USG images which also includes the comparative performance analysis of the proposed techniques with other types of existing machine learning techniques for PCOS detection.

Finally Chapter 6 of the thesis presents the discussion and conclusion of the research work with the thesis outcome, implications, limitations and future plan of this study.

CHAPTER 2

THEORETICAL BACKGROUND

This chapter briefly discusses the key concepts to provide basic theoretical knowledge regarding the background of this thesis. At first, a preliminary discussion on Polycystic Ovary Syndrome(PCOS) is presented. Then the automatic detection of PCOS has been elaborated which has been subdivided into two subsections: firstly, the existing studies of PCOS detection with machine learning techniques using patients' clinical symptom data; and secondly, PCOS detection with machine learning techniques using patients' USG image data. For both cases, the relevant research gaps of these domains have also been demonstrated.

2.1 Polycystic Ovary Syndrome(PCOS)

Polycystic Ovary Syndrome(PCOS) which affects almost 15% women globally in their reproductive age, is considered to be amongst the most prevalent endrocrinological disorders (Dapas and Dunaif 2022). It is usually driven by an excessive rise in male androgen hormones in the female body creating a long-term hormonal imbalance that has a detrimental influence on healthy ovarian function and results in the development of several cysts within the ovary (Abraham Gnanadass, Divakar Prabhu, and Valsala Gopalakrishnan 2021). This condition in women body shows heterogeneous symptoms that includes anovulation with irregular menstrual cycles, amenorrhea and oligomenorrhea ; hyperandrogenism with acne, hirsutism ; increasing rate of obesity and polycystic ovarian structure with numerous cysts inside ovaries (Rodriguez Paris et al. 2022). It is a complicated condition that impacts the reproductive, metabolic, and psychological systems in women body and is strongly linked to several other serious disorders (Joham and Teede 2022). PCOS is known as the primary cause of anovulatory infertility and is correlated to type 2 diabetes, thyroid problems, increased depression, sexual dissatisfaction, and other conditions which drastically reduces a healthier lifestyle of human being (Kałużna et al. 2020). Moreover, endometrial and ovarian cancer are seem to be more probable to occur in women with PCOS, each of which be catastrophic if not detected timely (Meczekalski et al. 2020).

PCOS has been linked to a number of disorders resulting in diverse symptoms in patients' bodies compared to normal ovulatory women, including type-2 diabetes, cardiovascular anomalies, hypertension, dyslipidemia, insulin resistance, increased Endometrium thickness and so on (Anagnostis, Tarlatzis, and Kauffman 2018; Jamil et al. 2015; Iatrakis et al. 2006). Furthermore, PCOS also causes a variation in the range of hormonal secretion such as luteinizing hormone (LH), Follicle-stimulating hormone (FSH), Anti-Müllerian Hormone (AMH) etc.(Garg and Tal 2016; Malini and K. R. George 2018). Additionally, some more indicators are identified to be strongly associated with PCOS including undesirable facial/-body hair, accelerated hair loss, dark spots on the skin, higher BMI, obesity and abdominal obesity with increased hip ratio, dietary habits with excessive fast food intake etc. (Usmani, R. Rehman, and Akhtar 2014; Couto Alves et al. 2017; S. George and Alex 2021). As a result, the standard clinical detection approach for PCOS is very critical, and also the accuracy as well as reliability of this anomaly identification and interpretations are heavily reliant on the physician's competence in this context.

The most prominent practice used by physicians to accurately diagnose PCOS is to examine ultrasound images of the ovaries to examine for the existence of multiple follicular cysts and hence determine whether or not the patient has PCOS (Pulluparambil and Bhat 2021). Ultrasound imaging is a well-known diagnostic technology that uses ultrasound waves generated by a transducer to create images of human body parts that provide real-time, precise anatomical and physiological information (Moran and Thomson 2020). In the case of ovarian ultrasound imaging, the ultrasonographers or radiologists visually assess the obtained ultrasound pictures for any abnormalities in the ovaries, and if several cysts with higher measurements of diameters are observed, then PCOS is diagnosed (Acharya, Molinari, et al. 2015).

However, by using suitable, symptom-focused, long-term, dynamic therapy and medication; the abnormalities in PCOS can be reversed; for which a timely and appropriate diagnosis of PCOS is crucial. A wide range of medical professionals frequently utilize the following three factors popularly known as ‘Rotterdam criteria’ to determine if a patient has polycystic ovarian syndrome (PCOS): menstrual irregularities, hyperandrogenism, and the presence of multiple cysts on ovarian ultrasound images (Rui Wang and Mol 2017). But, due to the heterogeneity of the associated symptoms of PCOS, it often becomes ambiguous to the clinicians to detect the condition at an early stage. Furthermore, the most accurate way to diagnose PCOS is to examine the existence of multiple cysts inside ovaries using ultrasonography(USG) imaging technology by expert radiologists (Bednarska and Siejka 2017). Since this technique also depends on the observer and these images include a lot of noise; clinical assessment, often becomes time-consuming and challenging, with a greater chance for human error. As a consequence, the expensive, device-reliant, and time-consuming procedure of diagnosing PCOS causes many young women who have this critical disorder to go undiagnosed and untreated for extended periods of time; worsening the disease condition.

2.2 Automatic Detection of PCOS

Number of studies have been conducted to investigate computer-assisted PCOS detection techniques, which offer substantial advantages such as rapid identification of the condition in the shortest time frame with the least amount of diagnostic error and human effort. (Isah, Usman, and Tekanyi 2015; M. N. Islam, S. R. Khan, et al. 2021). With the massive expansion of healthcare data and utilization of information technology, machine learning techniques are being one of the most widely used, efficient, and promising predictive strategies, which can analyze and retrieve key information from immense amounts of heterogeneous clinical data in order to detect diseases intelligently (Tchito Tchapga et al. 2021; Callahan and Shah 2017; M. N. Islam, Karim, et al. 2020). So far, PCOS detection has been conducted using two kinds of data modalities: Clinical-Symptom data and Ultrasound images of patients.

2.2.1 Detection with ML Techniques using Symptom Data

Researchers have applied various machine learning techniques in this context to detect PCOS condition from patient's symptom dataset. For example, to categorize between PCOS and non-PCOS criteria , Danaei et al. (Danaei Mehr and Polat 2022a) employed Extra Tree, Adaptive Boosting (AdaBoost), Bagging Ensemble with Random Forest and Multi-Layer Perceptron (MLP) classification models which were then evaluated through performance parameters using the reduced subgroups of features obtained by filter, embedded, and wrapper feature extraction techniques. Again, Boomidevi et al. (Boomidevi and Usha 2021a) suggested an artificial Neural network(ANN) model for detecting PCOS at an early stage where a comparative performance analysis had been conducted using different neural network optimizer to explore the best performing ANN design for classifying dataset into two classes: PCOS and Non-PCOS. Another related work in this field has been conducted by Prapty et al. (Prapty and Shitu 2020a), in which they investigated four different machine learning classifiers to categorize PCOS and non-PCOS records and compared their results where the Random Forest classifier outperformed the others; and then employing that Random Forest classifier a decision tree is developed to identify the top features responsible for PCOS. Denny et al. (Denny et al. 2019a) also proposed a framework named 'i-Hope' as a paradigm for early identification and prediction of PCOS based on optimum yet promising indicators; here they used a patient survey of 541 records to design the proposed framework, in which 8 potential features from diagnostic and metabolic test results were selected using SPSS and the Principal Component Analysis (PCA) method based on their importance, and then applied to seven types of traditional ML classifiers to find the best performing model.

Since PCOS is associated with a wide range of symptoms as features, a few studies have emphasized on employing various feature reduction approaches before using machine learning models to accelerate the training process (M. N. Islam, U. Hasan, et al. 2022; Suha and Sanam 2022b). For example, Inan et al. (M. S. K. Inan, Ulfath, et al. 2021b) suggested a strong sampling technique that includes both oversampling and undersampling procedures to boost minority samples; then applied two types of feature selection techniques : Chi-Square test for categorical and Analysis of Variance (ANOVA) test for numerical attribute selection; and then applied six types of machine learning cclassifiers where XGBoost classification model outperformed others. In another relevant article in this domain, Nandipati et al. (Nandipati, Ying, and Wah 2020a) used RFE-LR, RFI-ECT, SelectKBest/Chi2 and Forward Backward propagation techniques to find the top 10 and 24 features from all 42 features in the dataset, and then applied seven types of traditional ML classifiers in two different types of implementation platforms: Python-Scikit Learn package and RapidMiner; in addition, performance comparisons between various classfiers were assessed utlizing complete (40 features) and selected features (10 and 24 features) to find the best performing classifier. Munjal et al. (Munjal, Khandia, and Gautam 2020a) used a genetic algorithm and WEKA (Waikato Environment for Knowledge Analysis) software to identify the nine primary features associated in (PCOS) illness development and then utilized those reduced set of features over three types of ML classifiers in PyCaret platform to predict the disease using minimal attributes. While, for extracting the most significant attributes from a dataset comprising 26 attributes of 303 instances, Meena et al. (Meena, Manimekalai, and Rethinavalli 2015a) suggested an approach based on Neural Fuzzy Rough Set (NFRS) and Artificial Neural Network (ANN) techniques; and then applied those reduced set of features in four different types of classification models to detect PCOS where the performances enhanced in comparison to other five types of traditional feature selection methods.

However, from the earlier studies in this domain using symptom dataset it can be observed that, though there are number of contributions suggested from researchers across the world where several machine learning stretegies had been applied to detect PCOS; rarely has any researcher investigated the possibilities and efficacy of employing several types of ensemble machine learning techniques (bagging, boosting, and stacking) in this context. Ensemble techniques are a potential state-of-the-art solution for various machine learning difficulties, since they may considerably enhance a single model's forecasting performance by training multiple models and integrating their predictions (Sagi and Rokach 2018). Recently, several researchers have successfully employed ensemble machine learning strategies in other fields of healthcare predictions. For example, Jabbar et al. (Jabbar 2021) proposed an ensemble learning strategy to solve the challenge of breast cancer data categorization; Kaushik et. al.

(Kaushik et al. 2020) developed an ensemble of multi-headed ML architectures to forecast the average weekly expenditures on two pain drugs taken by patients etc.

From the previous related works it is also observable that, most of the studies have picked a specific reduced subset of features from the existing dataset through applying feature reduction techniques and then performed machine learning classification using that reduced feature subset. But, hardly any studies have explored multiple reduced feature subsets with different numbers and combinations of attributes from the complete dataset. Also, rarely they have investigated how the performances of various machine learning classification techniques might alter when different feature subsets with various combinations and numbers of attributes extracted from multiple feature reduction methods are being used. Furthermore, there has been minimal attention on investigating and validating whether the retrieved decreased features are genuinely important or not in terms of real-life clinical diagnosis of PCOS via a cross-check involving relevant healthcare specialists.

Thus, this research focuses on addressing these research gaps in this area with the goal of detecting PCOS more effectively and efficiently utilizing the optimum numbers of features. Therefore, a stacking ensemble classifier has been designed, trained, and evaluated as well as the performances of various forms of ensemble and conventional machine learning approaches have been investigated in this study, employing different sets of attributes acquired from feature selection methods.

2.2.2 Detection with ML Techniques using USG Image Data

Despite the fact that examining ultrasonography images is the most usual method to diagnose ovarian abnormality, the correctness and reliability of visual interpretations are frequently dependent on the observer's expertise and also several types of noises have an impact on ultrasound imaging making it harder for observers to diagnose (Acharya, Sree, et al. 2013). Therefore, to replace this arduous, error-prone, and time-consuming manual method of PCOS diagnosis; several researchers worldwide are exploring computer-assisted techniques of follicle identification and diagnosis of PCOS, which provide significant benefits such as rapid ultrasound image processing in the quickest time frame with reduced diagnostic mistake and human involvement (Isah, Usman, and Tekanyi 2015).

For developing computer-aided PCOS follicle detection system, the implementation of various forms of digital image processing techniques is the most frequently utilized strategy. For example, Mandal et al. (Mandal, D. Saha, and Sarkar 2021) suggested a technique to diagnose PCOS by automatically segmenting cysts and follicle regions from ultrasound images; for which they employed multiple digital image processing steps such as histogram equalization, K-means clustering, median filtering, and morphological erosion on 19 ultrasound pictures. Yilmaz et al. (YILMAZ and ÖZMEN 2020) tested and compared two methods of follicle detection using image processing techniques to diagnose PCOS; where the first approach includes noise reduction (Median, Average, Gaussian, and Wiener Filters), contrast modification (histogram equalization and adaptive thresholding), binarization, and morphological procedures.; and the second approach comprises of noise reduction (Gaussian Filter and Wavelet Transform), k-means clustering, hole filling and morphological operations. Further, they compared the analysis results using two performance metrics : False Acceptance Rate (FAR) and False Rejection Rate (FRR). Gopalakrishnan et al. (Gopalakrishnan and Iyapparaja 2020) have suggested an image processing-based strategy for PCOS detection utilizing the combination of modified Otsu method with active contour to determine the precise quantity of cysts from the ultrasound ovary image; their proposed method contains mainly two parts that is image pre-processing and follicle identification. In pre-processing stage they performed Region of Interest (ROI) extraction, speckle noise reduction using various strategies; then exploring the best performing filter technique they used the modified Otsu approach with active contour method to accomplish the second phase of follicle identification, which included image segmentation and feature extraction. Setiawati et al. (Setiawati, Tjokorda, et al. 2015) suggested a clustering approach that can be utilized for PCOS detection employing Particle Swarm Optimization (PSO) technique with a non-parametric fitness function to create more compacted and converging clusters for follicular segmentation.Sitheswaran et al.(Sitheswaran and Malarkhodi 2014) used object growing methodology to detect PCOS in two stages; where first step pre-processing comprised median filtering, local maximum extraction with ROI selection, and the second stage of follicle detection included cost map development, which finally provided a convex hull of probable follicles to decide object growing. Mehrotra et al. (Mehrotra, Chakraborty, et al. 2011) proposed another method where the input ultrasound picture would be preprocessed for noise reduction and contrast enhancement using a multiscale morphological technique and then the follicles would be segmented using scanline thresholding; they also conducted a comparative analysis of the manual result with the acquired result to assess its effectiveness. Again, Deng et al. (Deng, Y. Wang, and P. Chen 2008) suggested another automated scheme in which an adaptive morphological filter was used to filter the input ovary ultrasound picture, then for contour extraction a modified labeled watershed algorithm was employed and lastly, to detect PCOS, a clustering algorithm was used to locate anticipated follicular cysts. Thus, these investigations mostly used image processing to locate follicles in ovarian USG pictures, but they seldom classified the images into PCOS or non-PCOS criteria.

Several researchers combined machine learning models with digital image processing techniques to develop an automated machine learning based PCOS detection system which will not only detect follicles but also will classify them into PCOS or non-PCOS classes. For example, Rachana et al. (Rachana et al. 2021a) performed image enhancement, histogram equalization, Otsu thresholding, binarization, noise reduction, segmentation, feature extraction, and other relevant phases of image processing; and after that they used the KNN machine learning classification algorithm to categorize 50 USG photos. Nilofer et al. (Nilofer et al. 2021a) proposed a method where the images were first preprocessed using noise removal with median filter, segmentation using k-means clustering algorithm, feature extraction using GLCM method; and then a hybrid technique with artificial neural network (ANN) incorporating Improved Fruit Fly Optimization (IFFOA) was utilized to classify the images. On 65 ovary USG images, Gopalakrishnan et al. (Gopalakrishnan and Iyapparaja 2019) employed multiple conventional image processing approaches, including image enhancement, thresholding, noise reduction, Canny edge detection method to detect the follicle edge, and Scale-Invariant Feature Transform method to extract essential features; and furthermore they were classified using the Support Vector Machine (SVM), Decision Tree, and Naive Bayes classification algorithms, with SVM outperforming other models with 94.40% accuracy. Purnama et al. (Purnama et al. 2015a) suggested another approach for PCOS identification, in which USG images are first preprocessed through various phases to create binary images, which are then segmented using edge detection, labeling, and cropping; later, based on the feature vectors resulted from feature extraction step using Gabor wavelet, 3 types of classification algorithm was implemented with SVM-RBF kernel classification method achieving the highest accuracy of 82.55%. Deshpande et al. (Deshpande and Wakankar 2014) suggested another strategy, in which they used image processing techniques like contrast enhancement, filtering, feature extraction using Multiscale morphological approach and segmentation to determine the number of follicles in ovarian ultrasound pictures; following that, integrating the number of follicles acquired from the image processing stage with other features of that patient such as body mass index (BMI), hormone levels etc. they applied Support Vector Machine classification algorithm to detect PCOS.

Furthermor, Deep learning (DL) strategies have nowadays gained a lot of momentum in clinical diagnosis using medical images since they have the advantages over conventional machine learning approaches to bypass the feature extraction and traditional image processing steps as the neural network itself conduct these tasks (Brattain et al. 2018; Thapa et al. 2020; Yadav et al. 2022). A few researchers recently have also utilized deep learning methods for PCOS detection. For example, Vikas et al. (Vikas, Radhika, and Vineesha 2021) proposed a deep learning method for detecting PCOS, in which they initially implemented a three-layer Convolutional Neural Network (CNN) for detecting PCOS, then enhanced the model's accuracy by incorporating data augmentation method, and finally obtained the best accuracy by using transfer learning (VGG16 pretrained model) with fine-tuning. Cahyono et al. (Cahyono, Mubarok, Wisesty, et al. 2017) suggested a Convolutional Neural Network (CNN) having 6 layers to classify 40 PCOS and 14 non-PCOS USG images, with the feature extraction phase done automatically using deep learning.; they employed the softmax activation function and tested the model's performance using several alternative dropout and learning rates, with the best F1-score on test data being 76.36%.

Therefore, according to the previous studies PCOS and non-PCOS usg image identification has been so far conducted in two different ways: one group used image processing techniques to perform picture segmentation with feature extraction, and then a machine learning model to categorize the pre-processed images ; another group had applied deep learning methods to detect PCOS from USG images avoiding the image processing steps. Even though the PCOS detection had been conducted employing these techniques, but The follicular segmentation utilizing digital image processing requires cautious execution of several image processing stages. In addition, the extraction method using digital image processing techniques need to be very carefully tuned based on the varied picture quality and formats, which may need extra computing complexity and time, and thus each step becomes quite laborious. Also, it has been observed that employing digital image processing technique for feature extraction and then using conventional machine learning classifiers comparatively provides less accuracy in comparison to the deep learning methodologies. Also, a relatively small number of studies that used solely deep learning to diagnose PCOS encountered other difficulties, such as the prolonged processing times and high computational power necessary to handle the enormous amount of pictures needed to detect PCOS and thus it may become quite challenging to implement a practically usable interface for patients or medical professional using these deep learning approaches.

However, from the earlier studies it can be observed that, though there are number of contributions suggested from researchers across the world where several machine learning stretegies had been applied to detect PCOS; rarely has any researcher investigated the possibilities and efficacy of employing several types of ensemble machine learning techniques (bagging, boosting, and stacking) in this context. Ensemble techniques are a potential state-of-the-art solution for various machine learning difficulties, since they may considerably enhance a single model's forecasting performance by training multiple models and integrating their predictions (Sagi and Rokach 2018). Recently, several researchers have successfully employed ensemble machine learning strategies in other fields of healthcare predictions. For example, Jabbar et al. (Jabbar 2021) proposed an ensemble learning strategy to solve the challenge of breast cancer data categorization; Kaur et al. (R. Kaur, Doegar, and Upadhyaya 2022) applied stacking ensemble machine learning technique with an aim to identify brain tumor using Magnetic Resonance Imaging(MRI) of brain etc. Also, in a few recent research of other healthcare fields, the strengths of the CNN technique for feature extraction and the traditional machine learning technique for classification have been coupled to categorize medical images. For example, Pang et al. (Pang et al. 2019) developed a technique of protein subcellular localization for Alzheimer's disease prediction by using CNN as a feature extractor to automatically extract features from the original sequence data and XGBoost as a classifier to determine the subcellular localization based on the CNN's outcome. Such kind of hybrid approaches have rarely been employed in case of the area of PCOS detection using USG images. Thus, this study has proposed an extended and novel machine learning technique where CNN architecture has been employed as the feature extractor and then stacking ensemble machine learning has been used for classification with an aim to categorize PCOS and non-PCOS USG images.

CHAPTER 3

RELATED WORKS

This chapter discusses the systematic literature review(SLR) on the detection of Polycystic Ovary Syndrome (PCOS) with computer-aided techniques. In this chapter, firstly the methodology for the literature review has been discussed which has been subdivided into planning, execution and assimilation phases. After that, the data extraction and analysis has been demonstrated in an organized manner which includes the details of six types of data extraction themes that have been followed for conducting in this SLR. Following that the review findings from the data analysis have been reported and summarized. Basing on the review findings, a number of research gaps and future research opportunities are prescribed afterwards. Lastly, a critical summary is presented to emphasize the issues that motivated to conduct further study regarding the focused research opportunities.

3.1 Methodology for Literature Review

Literature reviews are generally the initial stage of any research work that provide published knowledge and reliable findings from relevant research on that study topic (M. N. Islam, I. Islam, et al. 2020). Systematic Literature Reviews(SLR) offer a more comprehensive, well-organized summary that follows accepted practices and also outline, assess, and discover the research gaps of the published studies being unbiased in accordance with some study objectives and research questions (I. Islam and M. N. Islam 2022). From methodological perspective, the SLR in this study has been conducted in three phases following the approaches suggested by Kitchenham et al. (Kitchenham 2004; Kitchenham and Charters 2007) and Tandon et al. (Tandon et al. 2020). The outline for the SLR has been shown

Table 3.1: Inclusion and exclusion criteria for SLR

Inclusion Criteria	Exclusion Criteria
Articles with a specific focus on automated computer assisted techniques for PCOS detection	Articles that mentions PCOS but do not includes computer assisted techniques for PCOS detection
Articles that are available in complete or full-text	Articles published as short surveys, reports, assumptions, notes.
Articles published since 2005 to 2022	Duplicate research articles published in multiple repositories
Research articles published in peer-reviewed journals, conference proceedings or as thesis dissertations	Studies that focus only on clinical diagnosis aspects for PCOS detection
Articles written in English language	Research articles conflicting the objectives of the systematic review

in Figure 3.1.

Fig. 3.1. Outline for systematic literature review

3.1.1 Phase 1: Planning

The first stage of the study is planning that includes the preparation before executing the review with requirement analysis for conducting the SLR; identifying the research objectives and research questions and then formulating the protocol for conducting SLR.

3.1.2 Requirement Analysis:

At the requirement analysis phase, the need of performing this SLR has been investigated and therefore necessary background knowledge has been gathered including clinical perspectives of PCOS as well as how computer assisted techniques can be beneficial in PCOS

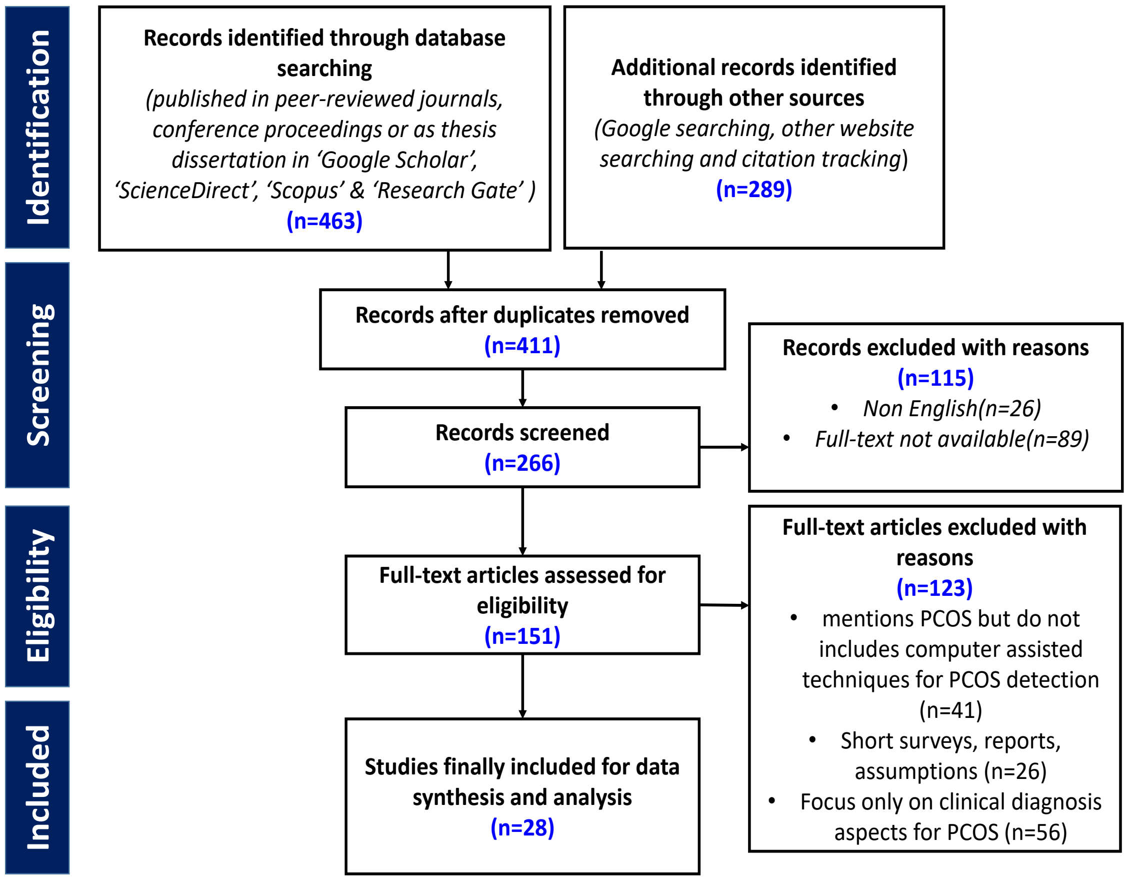


Fig. 3.2. PRISMA flow diagram for selection of articles

detection (Suha, M. N. Islam, et al. 2022). At this stage, it has been observed that numerous research have been carried out that demonstrate how effectively PCOS may be recognized at an early phase through various ML-based techniques using patients' records (Suha and Sanam 2022c). However, hardly any study has been carried out where the summary of all these studies have been presented or the findings have been investigated to explore the limitations and future aspects in this research domain. As a result, the requirement analysis step amply demonstrates the need for an SLR in this area.

3.1.2.1 Research Objectives Identification:

The core objective of the research is to explore the effectiveness and significance of computer assisted techniques for detecting PCOS; and summarize the findings, limitations and future scopes of the existing studies in a systematic manner. Thus, at this stage of planning the following auxiliary goals have been identified that can help to achieve the primary goal of the study:

- RO1: to analyze the purpose, goals, and publication histories of the existing research in this domain;
- RO2: to investigate the type of data employed for PCOS identification in relevance to the research's scope;
- RO3: to explore different types of computer assisted techniques and evaluation strategies that have been used for autonomous PCOS detection;
- RO4: to investigate the research findings and challenges to detect PCOS
- RO5: lastly, to structure potential future research scopes in autonomous detection of PCOS.

3.1.2.2 Research Protocol Formulation:

An initial research protocol has been formulated in this stage with inclusion and exclusion criteria for selecting the articles for reviewing. The inclusion and exclusion criteria has been shown in Table 3.1.

3.1.3 Phase 2: Execution

The execution part of the SLR is where the review has been conducted that involved searching, filtering, and selecting the study materials, after which the required data were extracted, assessed, and synthesized to explore the review outcomes. The activities of execution phase had been performed from February 2022 to July 2022.

3.1.3.1 Searching, screening and finalizing review materials

For selecting the studies to perform SLR, a systematic guideline known as 'Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA)' (Selçuk 2019) has been adapted in this study. The PRISMA flow diagram for selecting the review materials has been shown in Figure 3.2. The process consists of four phases: identification, screening, eligibility and included.

Table 3.2: List of keywords ulitilized for searching in database

SL No.	Keywords
1	”Polycyctic Ovary Syndrome (PCOS)”
2	”PCOS detection or diagnosis”
3	”PCOS detection with computer assisted techniques”
4	”Automated PCOS detection techniques”
5	”early screening of PCOS with machine learning techniques”
6	”Feature selection and prediction of PCOS”
7	”Automated Follicle identification in PCOS”

At identification phase the articles that have been published in peer-reviewed journals, conference proceedings or as thesis dissertation are searched from different familiar databases like Google scholar, Science Direct, Scopus, Research Gate etc. as well as in some additional sources like website searching, citation tracking etc. using some keywords. The list of keywords that have been used for identification phase are listed in Table 3.2. Initially total 752 research articles have been found in this way which are then minimized in the screening phase. At screening phase, duplicate records, articles with non English language and the articles that are not available in full-text are eradicated for reviewing that resulted into 151 articles. The next phase is eligibility, where the final article with are eligible for the SLR are selected meticulously considering all the inclusion and exclusion criteria. Therefore, after conducting all these selecting phases finally 28 articles are included for data synthesis and detail analysis of this study.

3.1.3.2 Data extraction, analysis and synthesize:

At this phase, data have been extracted, analyzed and synthesized from the selected research articles according to the data extraction theme shown in Figure 3.3. Here, six aspects of data extraction themes have been considered to fulfil the research objectives of the SLR. A list of queries are outlined on each theme in order to get the particular type of information needed.

- **Topological Association:** To examine the basic correlation and relevancy between the research literature, this theme investigates two queries:
  - Are the titles of the articles closely linked?
  - Are the keywords in the articles frequent?

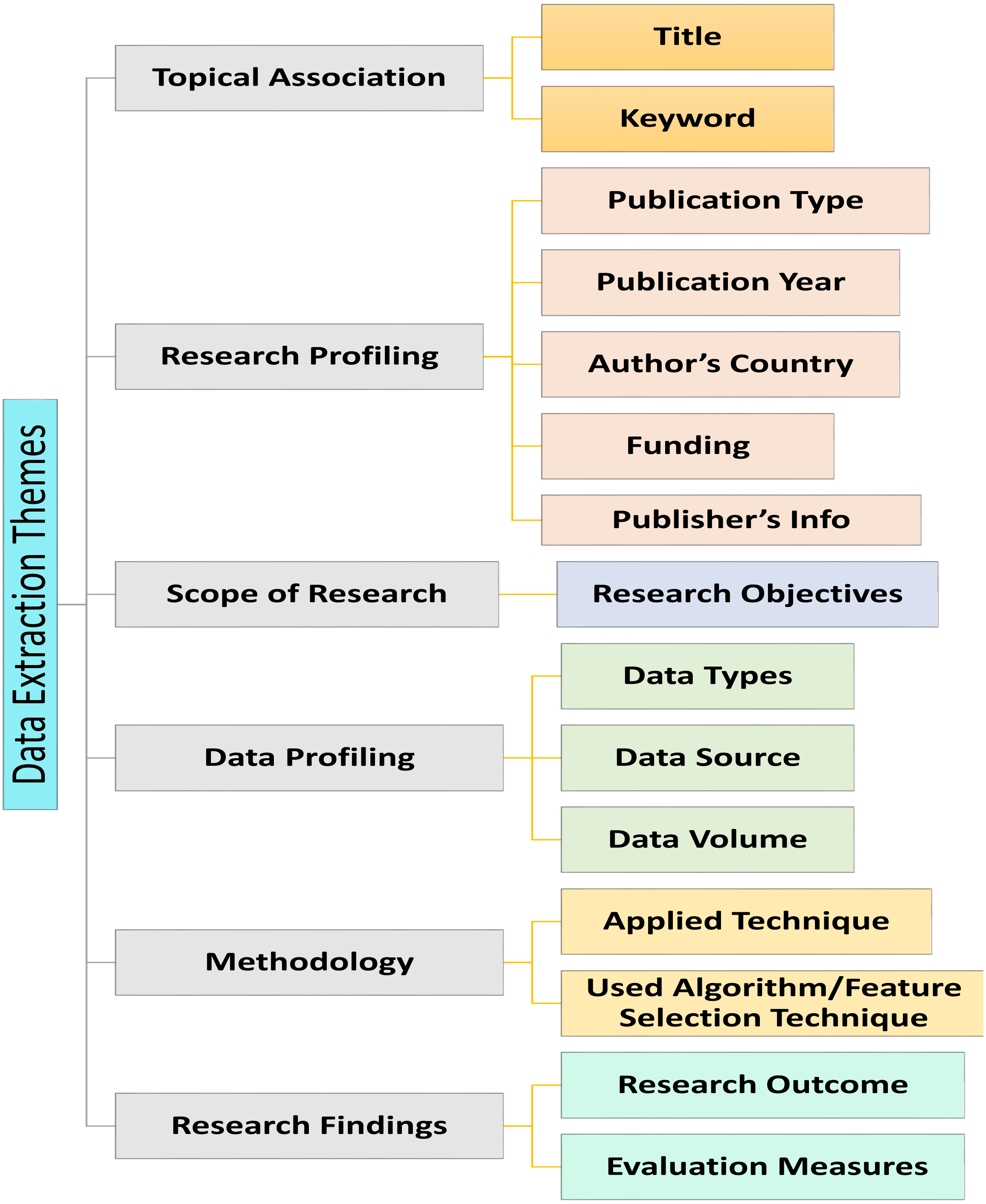


Fig. 3.3. Data extraction theme for systematic review

- **Research Profiling:** This theme depicts the research publication profiles of the selected articles employing four subqueries:

  – What are the publication types of selected studies ?

  – Which year did the studies got published?

  – Which countries are the authors of the articles from?

  - Was funding granted to conduct the research or not?
  - Which publishers published these research articles?
- **Scope of Research:** This theme focus on the scope and context of the research using the following query:
  - What are the research objectives of the selected studies?
- **Data Profiling:** To investigate the data utilized for the articles according to the research objectives, this theme utilize three queries:
  - What type of data are employed to predict PCOS?
  - What are the data sources that have been utilized in the studies?
  - What are the size or volume of the used dataset?
- **Methodology:** This theme focus on the research methodology that have been followed to meet the objective of the research employing two questions:
  - What kinds of techniques or technologies have been applied for predicting PCOS in the chosen articles?
  - What are the algorithms or feature selection/extraction techniques that have been used to implement the proposed methodology?
- **Research Findings:** Through this data extraction theme, the key research findings from the articles are structured employing the following queries:
  - what are major the research outcomes from the articles?
  - What evaluation measures have been used to assess the proposed methodology?

3.1.4 Phase 3: Assimilation

After conducting the execution step of the SLR, the final stage is assimilation that wraps up the review by reporting the key factors. At this phase, the research results from the selected articles are summarized to present the major findings. Following that, the possible research gaps, limitations and challenges from these investigations are listed. And finally, according to the analysis reports of all the findings the promising future research agenda are explored for autonomous and efficient detection of PCOS employing machine learning and computer assisted advanced methodologies.

3.2 Data Extraction and Analysis

The detailed description of data extraction and analysis using six themes has been discussed in this section which are conducted after the selection of the 28 eligible articles to perform SLR.

Fig. 3.4. Wordcloud of the Titles from selected articles

3.2.1 Topological Association

To examine the topological association in terms of the correlation and relevance of the selected articles, ‘Wordclouds’ have been generated in this study which is considered to be an appealing visualization technique for representing any text analytics quickly (Heimerl et al. 2014). Here, two different word clouds are created, one utilizing all of the article titles (see Figure 3.4) and the other using the keywords (see Figure 3.5); with an aim to address the two queries under the theme of topological association. The word cloud displays the frequency of the words in proportion to dimensionality that are included in the set of keywords and titles of the articles under evaluation; where the biggest and boldest words tend to be more significant and frequent compared to the smaller, less concentrated words. From the Figure 3.4, it is apparent that 'polycystic', 'syndrome', 'ovary', 'learning', 'ultrasound', 'detection', 'machine', 'using' etc. are the most frequently utilized words in the titles of the articles. Also, from the Figure 3.5, it is visible that the most frequent words utilized in the keywords of the articles are 'polycystic', 'PCOS', 'ovary', 'syndrome', 'ultrasound', 'follicle', 'feature', 'machine', 'classifier', 'segmentation' etc. Smaller-sized words from the clouds are not frequently utlized in the articles and may have been used as titles or keywords in some individual articles.

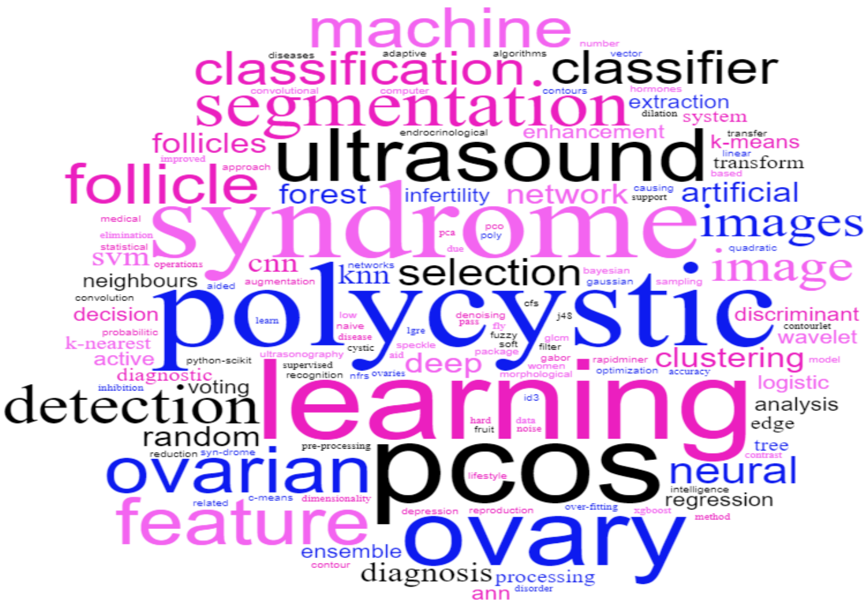


Fig. 3.5. Word cloud of the Keywords from selected articles

3.2.2 Research Profiling

The research profiling of the selected studies have been conducted in five perspectives: publication type, publication year, researcher's country , publisher's name and funding. The publication type includes the information that the selected articles are what type of research articles. From the pie-chart shown in Figure 3.6, it is observable that the selected articles are either from any journal or from conference proceedings. Here, in 28 selected articles 15 are from journals that covers 52% and 13 are from conference proceedings covering 48% of the publication type.

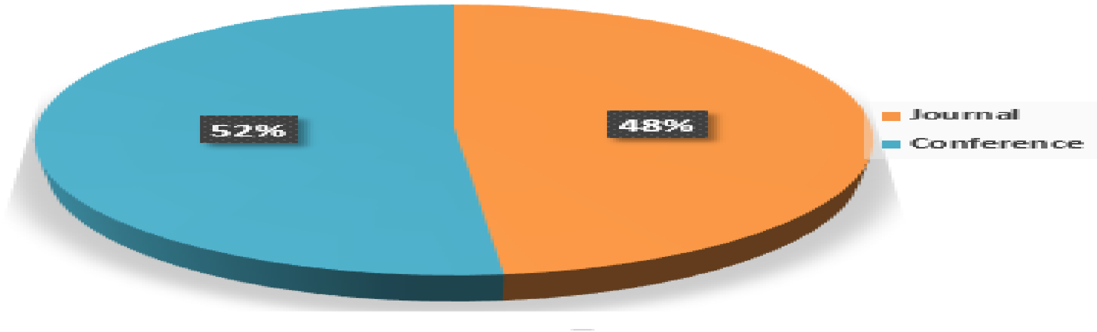


Fig. 3.6. Publication Type

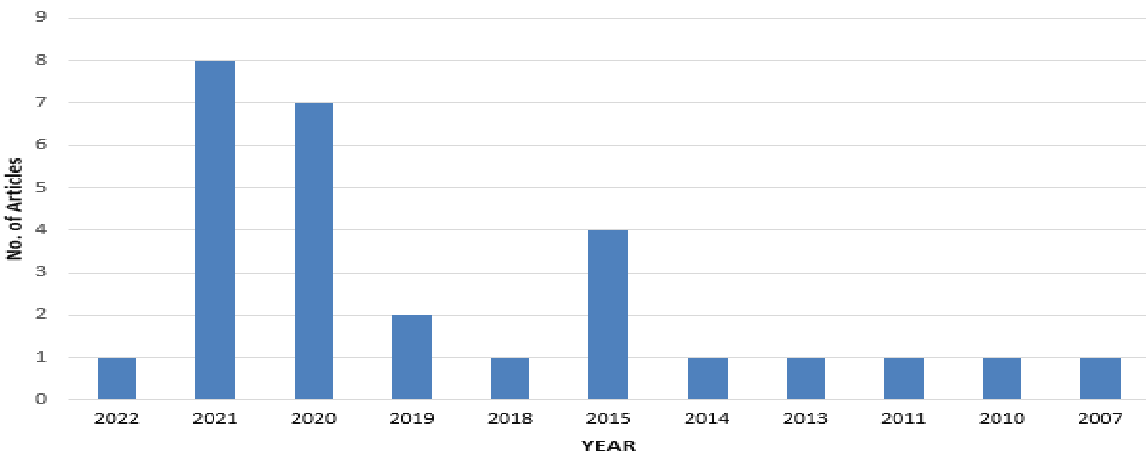


Fig. 3.7. Publication trend between year 2005 to 2022

The chosen research articles of this study have all been published in between the year 2005 to 2022 as mentioned in inclusion criteria. Figure 3.7 shows the publication trend in between this time frame, which reveals that most of the studies have been conducted in the year 2021. It is to be mentioned that, numbers of studies conducted in 2022 could not be included in this study as they were not available in full-text online by July 2022. However, from the Figure 3.7 it is also observable that, less number of studies had been conducted in this field before 2010 and then the trend of research works is increasing gradually in this domain. For addressing the countries from where the researches in this domain had been conducted, Figure 3.8 has been graphically represented. The name of the nation denotes the area of the university that the study's first author is a member of. From the figure it is noticeable that, most of the research works in this domain have been conducted in South-Asian subcontinent where the highest number of studies are from India (17) and then from Bangladesh (4). Other works which have been considered to review in this study are from Indonesia, Turkey, Malaysia, Lebanon and Canada.

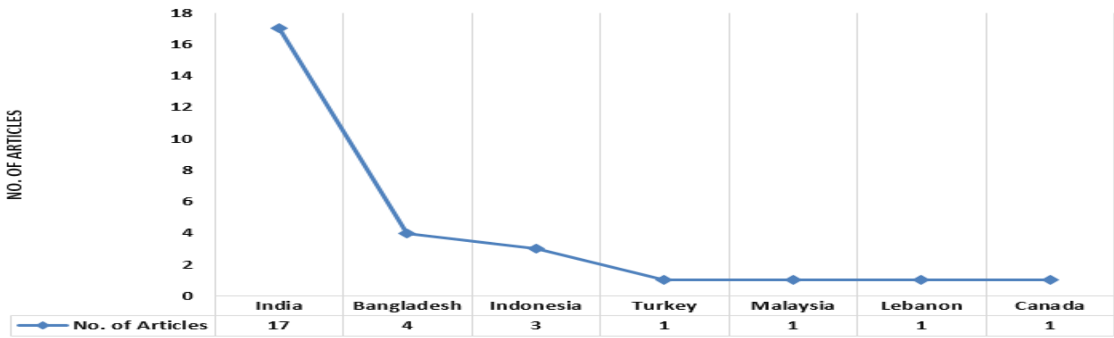


Fig. 3.8. Number of articles per country where the first author is affiliated

Under the data extraction theme based on research profiling, the following concern is to structure the articles based on its funding options. The donut chart illustrated in Figure 3.9 shows that, only 21% (6 out of 28) research works conducted in this domain were funded financially from any organization and most of the studies (79%) were not provided any financial supports to conduct the work. For addressing the final query under this theme, the publishers names are structured as per the number of articles in Figure 3.10. From the graphical representation, it is apparent that the highest number of studies that have been conducted to predict PCOS using computer-assisted techniques have been published by IEEE(11 articles), and then by Springer (6 articles). Several other publishers like IOP, Elsevier, IJSR, AMCI etc. have also published number of studies in this research topic.

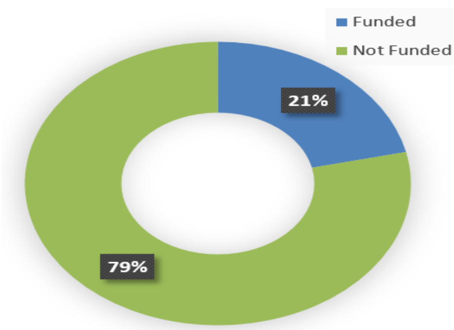


Fig. 3.9. Funding information

Table 3.3: Exploring research scopes of the selected articles

Category	Research Aim	Reference	Freq
1	To explore the most significant features and then detect PCOS from data-driven patient records	Danaei Mehr and Polat 2022b; Nandipati, Ying, and Wah 2020b; Munjal, Khandia, and Gautam 2020b; Prapty and Shitu 2020b; Tanwani 2020; Bharati, Podder, M. Mondal, et al. 2022a; Mehrotra, Chatterjee, et al. 2011; M. S. K. Inan, Ulfath, et al. 2021a; Meena, Manimekalai, and Rethinavalli 2015b; Denny et al. 2019b; Adla et al. 2021	11 (39%)
2	To detect PCOS from data-driven patient records	Nabi et al. 2021; Boomidevi and Usha 2021b; Hassan and Mirza 2020	3 (11%)
3	To propose an autonomous technique for classifying ovarian ultrasound images in PCOS or non-PCOS classes	B, Yalavarthi, and K 2021	1 (3%)
4	To propose a technique for follicle or cyst segmentation; and then classification into PCOS or non-PCOS classes using ovarian ultrasound images	Nilofer et al. 2021b; Dewi, Wisesty, et al. 2018; Padmapriya and Kesavamurthy 2015; Lawrence et al. 2007; Gopalakrishnan and Iyapparaja 2021; Purnama et al. 2015b; Hartati, Musdholifah, et al. 2019; Rachana et al. 2021b	8 (29%)
5	To detect the follicle or cyst in ovarian ultrasound images	Raj 2013; Hiremath and Tegnoor 2010; Kiruthika and Ramya 2014	3 (11%)
6	To classify the type of cyst or follicle using ovarian ultrasound images	Sumathi et al. 2021; Pathak and Kulkarni 2015	2 (7%)

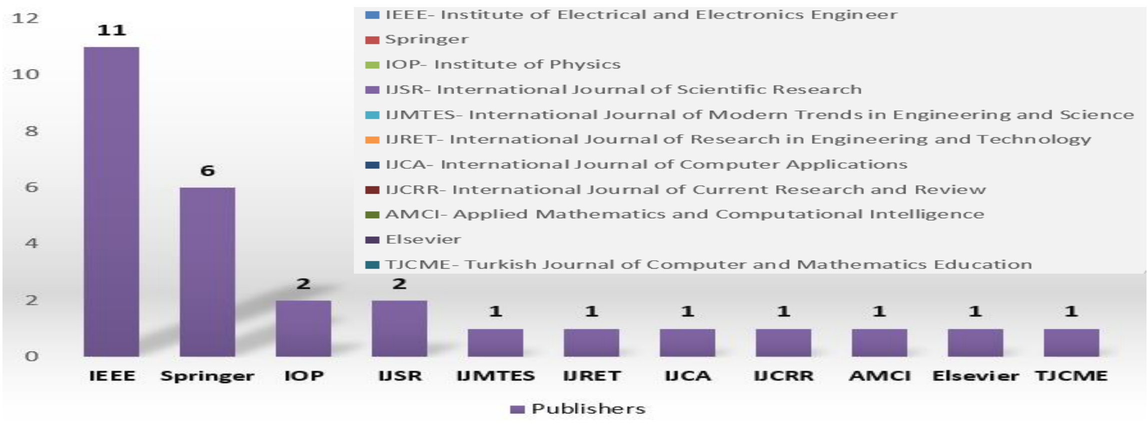


Fig. 3.10. Number of articles as per Publisher's name

3.2.3 Scope of Research

With an aim to explore the scopes of the selected research works, Table 3.3 lists the objectives of the studies as well as their frequencies.

From the Table 3.3, it is observable that the total 28 studies altogether have been divided into six categories based on their primary study objectives. The first group of studies are those which applied some techniques for feature selection to explore the most significant features from patient records and then using those reduced set of features they applied machine learning classification models to detect PCOS. In this SLR, most of the studies (39%) have been found with this research objective. For example, Danaei et. al (Danaei Mehr and Polat 2022b) proposed a technique for PCOS diagnosis where the most important 33,30 and 28 features were selected using several feature selection technique from 42 features and then using those set of features numbers of machine learning classification algorithms were trained and tested to classify among PCOS and non-PCOS patients. However, another group of studies (for example the work in (Nabi et al. 2021)) categorized as second types of research works in Table 3.3, did not conduct any feature selection tasks; rather they just applied machine learning classification models employing all the attributes for PCOS detection using data-driven patient records. A few number of studies have been explored (11%) with such kind of research aim.

Again, the third category of studies in this domain as per the research objective had proposed machine learning based autonomous technique to classify PCOS or non-PCOS patients using ovarian ultrasound images. As only one existing work in (B, Yalavarthi, and K 2021) has been found to be reviewed in this SLR, this category is relatively less explored. Following that, the fourth category of research works conducted studies with an aim to initially detect or segment the cysts using various techniques from ovarian ultrasound images; and then employing the extracted features of detected cysts apply machine learning classification methods to classify images into PCOS or non-PCOS category (for example the work conducted by Nilofer et.al (Nilofer et al. 2021b)). Comparatively a lot of studies (29%) have been conducted with this research aim. Moreover, another small group of studies (11%), mentioned as the category 5 in Table 3.3, had carried out their research (for example as performed by Raj et al. (Raj 2013)) with the goal of locating the follicles or cysts as the region of interest in ovarian ultrasound images of PCOS patients. Finally, the last category of research works (7%) had performed their studies with an aim to classify the type of the detected cysts from the ovarian ultrasound images of PCOS patients as done by Sumathi et al. (Sumathi et al. 2021) and Pathak et al. (Pathak and Kulkarni 2015). Thus, from this analysis it is evident that, the researchers have explored the detection of PCOS using computational intelligence in various aspects and aims.

3.2.4 Data Profiling

This data extraction theme focuses on exploring the data profiles of the research article as well as providing a relationship between the research aims and data types of the studies. Table A.1 shows the data profiling summary of the selected research articles. The table provides information about the data type, source and volume of the articles and their frequencies as well as the table also divides them according to the research aim category.
From the table, it is observable that mainly two kinds of data have been utilized to predict PCOS with computer-assisted techniques. The 14 studies with research objective category 1 and 2 (see Table 3.3) have used patient records as data, where various patient history, symptoms and diagnosis results are represented numerically or categorically as features in the dataset. On the other hand, the other 14 studies with research objective 3,4,5,6 (see Table 3.3) have employed ovarian ultrasolonography (USG) images from patients as dataset to predict PCOS. The highest number of studies (10 articles) have employed an open source

data from Kaggle repository Vedpathak 2020, which consists of the records of 541 patients having 44 attributes collected from various hospitals of India. Other studies have collected data from various hospitals or infertility centers as mentioned in table A.1. However, 9 publications have been discovered which just mentioned anonymous online open source repositories without expressly disclosing their data sources for USG image data.

3.2.5 Methodology

Table 3.4: Applied Techniques for the selected studies

Name of Applied Technology/Technique	Ref	Freq
Machine Learning classification	Danaei Mehr and Polat 2022b-Rachana et al. 2021b, Sumathi et al. 2021, Pathak and Kulkarni 2015	25
Deep Learning Techniques	Boomidevi and Usha 2021b; Nilofer et al. 2021b; Dewi, Wisesty, et al. 2018; B, Yalavarthi, and K 2021	3
Digital Image Processing Techniques	Nilofer et al. 2021b-Pathak and Kulkarni 2015	13
Feature Selection Techniques	Danaei Mehr and Polat 2022b-Adla et al. 2021	11

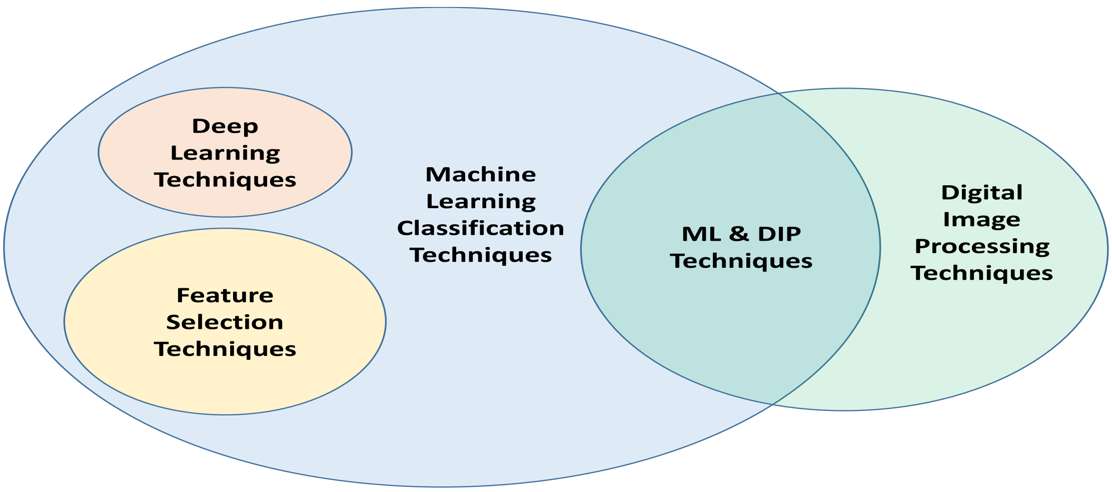


Fig. 3.11. Ven Diagram of applied research methodologies

This data extraction theme focuses on the research methodologies which have been employed in the the articles with an aim to autonomously detect PCOS. Table 3.4 shows the list of major techniques or technologies that have been applied in the articles to achieve the goal. From the table, it is noticeable that among the 28 selected articles, 25 studies have applied machine learning classifcation techniques. Among them, 3 articles have applied deep learning based methodologies.Various kinds of digital Image processing techniques have been applied by 13 articles which utlized USG images for PCOS identification. Moreover, 11 articles have been explored which applied various feature selection techniques to identify the most significant attributes among the attributes in the dataset. The list of several types of feature selection techniques that had been applied in these articles are provided in Table A.3. Furthermore, the core applied methodologies have been represented through a Ven-diagram in Figure 3.11 which depicts that, most of the articles here applied machine learning(ML) techniques and among them a few applied deep learning techniques or feature selection techniques. Moreover, there are some articles which applied digital image processing(DIP) techniques and a few have applied both ML and DIP techniques to achieve the research objectives.

Most of the articles (n=25) utilized different types of machine learning algorithms for predicting the presence of PCOS employing either the ovarian USG images of patients or their metabolic and clinical attributes. The list of ML algorithms have been listed in Table A.2. From the table it is observable that most of the articles utilized Support Vector Machine (n=12), Random Forest (n=10) and K-Nearest Neighbour (n=10) classification models for training. Different types of ensemble classifiers had also been employed by some articles, for example, voting hard and soft ensemble classifier, adaptive boosting, gradient boosting, eXtreme gradient booting, Categorical Boosting classifier etc. Limited number of studies had applied neural network based classification models such as Multi-layer perceptron, Artificial Neural Network, Convolutional Neural Network etc.

Different types of platforms, programming languages and integrated development environment (IDE) have been utilized to implement the research methodologies. For example, Munjal et al. Munjal, Khandia, and Gautam 2020b had implemented the feature selection applying genetic algorithm using Waikato Environment for Knowledge Analysis (WEkA) software. Nandipati et al. (Nandipati, Ying, and Wah 2020b) performed the same methodology in two types of application platforms: Pyhon-Scikit learn and Rapid Miner. However, most of the studies had implemented their proposed techniques with machine learning employing Python programming language based IDE such as, Pycharm, Pydev, Visual Studio, Google Colaboratory etc. Also the articles which have applied digital image processing techniques for follicle segmentation have mostly utilized MATLAB platform for implementation.

3.2.6 Research Findings

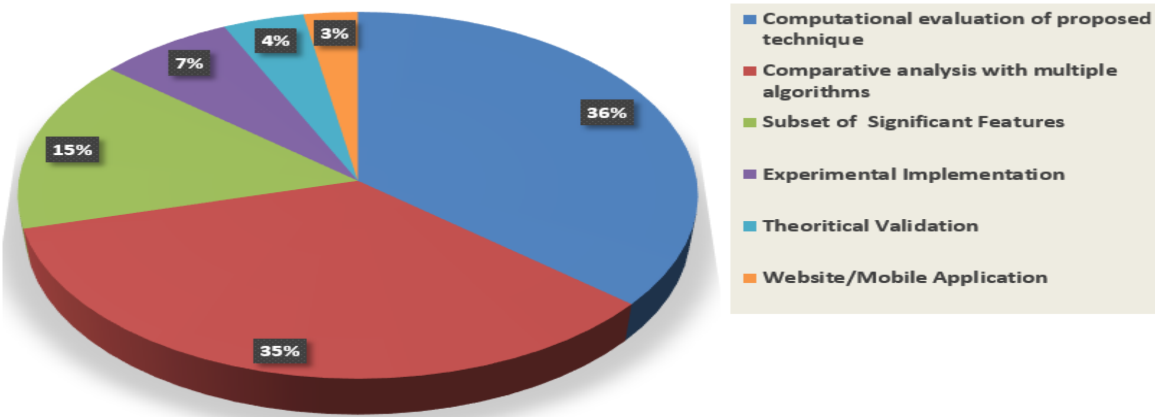


Fig. 3.12. Various type of research outcomes

The selected studies of this SLR provided several types of research outcome, though all of them had the key goal to detect PCOS through computer-assisted techniques. The research' findings can be summarized and divided into six groups: subset if significant features for PCOS identification, computational evaluation and validation of the proposed technique, comparative analysis with multiple algorithms or existing methodologies, architectural framework, theoretical validation and website/mobile application. Most of the articles (n=26) provided a research outcome with computational evaluation and validation of the proposed techniques, where they initially proposed a technique for PCOS detection and then evaluated and validated it through train and test dataset using various performance metrics. For example, Rachana et al.(Rachana et al. 2021b) proposed a technique based on image processing and machine learning classification models for PCOS detection; and then conducted a computational evaluation through train and test dataset of ovarian USG images with an aim to explore the efficacy of the proposed methodology. Moreover, a large group of articles (n=25) provided a comparative analysis between various types of machine learning classifiers with an aim to explore the best performing one or showed a comparison between the proposed computational technique and the manual process of PCOS detection. For example, Gopalakrishnan et al. (Gopalakrishnan and Iyapparaja 2021) provided their study findings with a comparative analysis of the manual identification vs. the proposed automated segmentation of follicles for PCOS detection as well as evaluated the error rates. Moreover, a subset of the most significant features for PCOS detection were suggested as study findings by certain research (n=11) using the dataset of patients' varied metabolic and clinical symptoms. For example, Danaei et al. (Danaei Mehr and Polat 2022b) applied various types of feature selection techniques such as filter methods, wrapper methods and embedded methods to explore different sets of features selected by different kinds of techniques; and finally suggested a set of features as research outcome which provides best accuracy while applying machine learning models. Another group of studies (n=5) offered experimental implementation of their proposed methods for PCOS detection or follicle segmentation from ovarian USG images. For example, Lawrence et al. (Lawrence et al. 2007) conducted an experimental impelemtation of their proposed methodology over 70 patients of Women's Health Imaging Research Laboratory (WHIRL) situated in Saskatoon, Canada where they validated the segmentation of the follicles in the ovarian USG images through both expert identification and automated detection process.

A few studies (n=3) provided a model for web-based or mobile application as research outcome to detect PCOS automatically using machine learning based techniques. For example, Denny et al. (Denny et al. 2019b) suggested an autonomous application model for PCOS diagnosis aid named 'i-Hope' that would be developed based on the attributes selected by Principal Component Analysis(PCA) technique and Random Forest Classification model. Moreover, some studies (n=2) proposed a technique for PCOS detection and validated it through theoretical explanation. An example of such kind of research outcome can be observed in the work conducted by Hartati et al. (Hartati, Musdholifah, et al. 2019) where they provided a framework with theoretical and mathematical validation for PCOS identification using ovarian USG images and the researchers hoped to use the suggested method in practice in the future.

However, for evaluating the performances of the proposed techniques, the articles applied different types of performance metrics. The list of the performance parameters with the references and frequencies are presented in Table A.4. From the table, it is observable that, most of the articles (n=25) employed accuracy to measure the efficacy of the proposed technique in result analysis. Precision, recall, F1-scores, specificity, confusion matrix are also utilized by a vast group of articles. Additionally, as PCOS detection is a practical challenge in the field of healthcare analytics, some publications (n=5) compared the reliability of autonomous detection with the manual identification of PCOS through expert clinicians. Other than these, Area Under the Curve(AUC) score, Receiver Operating Characteristic (ROC) curve, loss calculation, percentage of error rate, positive predictive value (PPV) calculation are also employed to measure the performances of the proposed techniques in few other articles, each with a frequency of 1 or 2.

3.3 Reporting the Review

3.3.1 Summarizing Review Findings

This SLR has explored 28 relevant articles which are all been conducted with an aim to detect PCOS in female body autonomously with computer-assisted techniques. For analyzing the articles in a systematic way, the data synthesis has been performed using six data exploration themes. The findings from each of the themes can be briefly summarized hereafter:

- In the first theme, the word clouds under topological association reveal that, the titles and keywords of the chosen studies are closely associated with each other with several similar and frequently utilized words where majority of the articles clearly mentions PCOS detection techniques in their titles or keywords.
- The research profiling of the the articles clearly indicates that, this area of research has gained almost equivalent publication interest in both journals and conferences from various publishers. The year-wise publication statistics also shows that the interest of the researchers in this domain is increasing day by day. Thus, this analysis indicates that there is a growing emergence with increasing interest by the researchers

worldwide in autonomous prediction of PCOS using computer-assisted techniques and machine learning. However, the authors origin statistics indicate that, most of the studies have been conducted in South-Asian subcontinent and majority of them are from India. Also, a few research have acquired funding to conduct such kind of research works.

- Investigating the scopes of the selected research articles, it has been explored that, though detecting PCOS using computer-assisted methods is the main goal of all the publications, each one also has certain distinctive research goals. The research objectives of all the articles have been categorized into six groups where 39% studies have the aim to explore the most important features for PCOS detection and then to detect PCOS based on those attributes. And, another large group of studies with 29% articles have the objective to conduct a segmentation of the cysts from ovarian USG images and then classify them into PCOS or non-PCOS criteria.

- The data profiling analysis of the articles indicates that, there are mainly two types of data that have been utilized for PCOS detection: one of which is patients' various symptoms with metabolic and clinical test reports represented as features of the dataset and another type of data is ovarian USG images of patients. However, most of the articles have used dataset which contains the records of Indian women collected from various hospitals or open source repositories.

- The research methodologies of the articles indicates that, there are mainly four types of technologies that have been applied in this domain: machine learning, digital image processing, feature selection techniques and deep learning. Here, most of the studies applied various types of machine learning classification models such as, random forest, KNN, SVM, DT etc. (see Table A.2). Also, various types of feature selection techniques have been employed by the articles to explore the most significant features for PCOS detection. A group of articles applied both digital image processing techniques for follicle segmentation and then applied machine learning techniques for PCOS and non-PCOS patient classification. However, only three articles have been found that utilize deep neural networks in this domain, limiting the exploration of deep learning approaches.

- Finally, the findings of selected articles have been grouped into mainly 6 categories where it has been observed that, most of the articles provided computational evaluation and validation as well as comparative analysis of the proposed techniques with other methods as research outcome. Various studies have utilised different types of performance metrics to evaluate the efficacy of their suggested methods. It has also been noticed that, though similar types of classifiers have been employed in multiple articles, each of them gained different performances because of their non-identical methods of implementation and data pre-processing techniques. For example, using the same dataset Danaei et al. (Danaei Mehr and Polat 2022b) acquired 98.89% accuracy whereas Munjal et al. (Munjal, Khandia, and Gautam 2020b) gained 85% accuracy both applying Random Forest classification model for PCOS identification.

3.3.2 Research Gaps and Future Research Recommendations

In the field of PCOS diagnosis utilizing computer assisted techniques, the results of this SLR highlight potential research gaps or constraints that still need to be addressed and also suggests possible future research areas for more effective methodologies. Following are some suggested future study initiatives that might be taken into account to fill in the research gaps. The possible research agenda are briefly outlined in Figure 3.13.

3.3.2.1 Exploring the integration of emerging advanced Techniques

The existing ML-based strategies for PCOS identification can be strengthened by incorporating a variety of emerging and advanced techniques. For example, ‘explainable AI or XAI’ can be integrated for PCOS detection which contains the potential to explain AI-based decision-making in clear terms to individuals and to a wider range of end users for making it easily interpretable (Y. Zhang, Weng, and Lund 2022). XAI is gaining much attention in the healthcare analytics as a great tool for clinicians to spread the use of ML-based procedures in practical fields. Moreover, ‘Federated Learning or FL’ is another cutting-edge

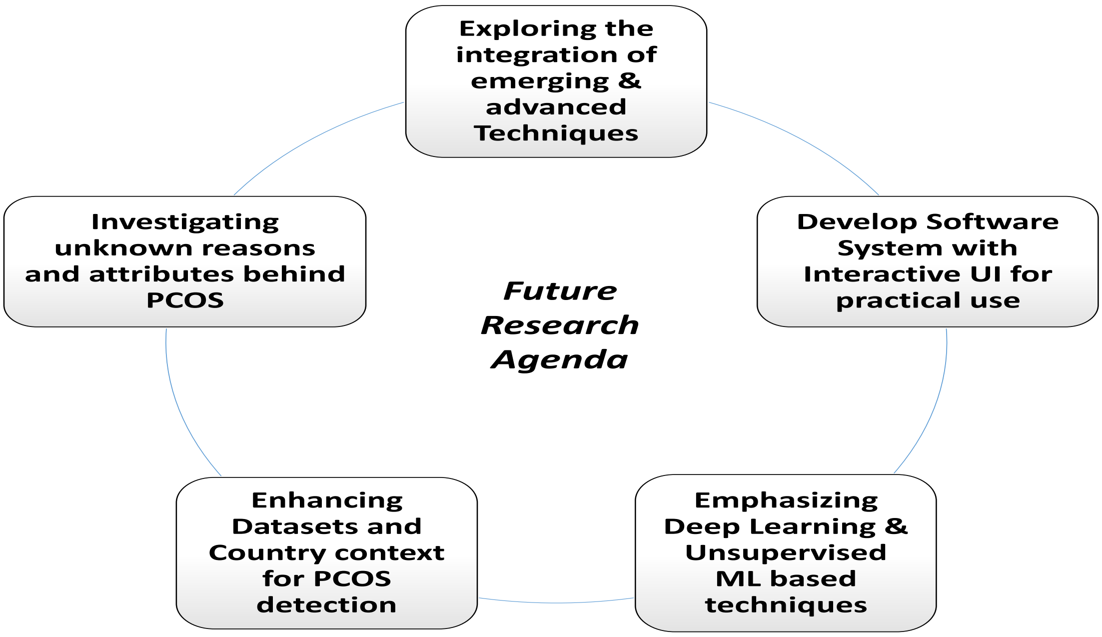


Fig. 3.13. Outline for future research agenda

method that can be incorporated to detection process of PCOS. FL uses the cloud to prevent security breaches by allowing different data owners to work together to train a model locally employing their own data without sharing data with anyone else (Md. Mahbubur Rahman et al. 2022b). Employing FL techniques in PCOS detection, several hospitals around the world can participate in developing a robust ML-based model for PCOS prediction without even disclosing their own datasets to others.

3.3.2.2 Develop Software System with Interactive UI for practical use

The existing studies in this domain have hardly focused on developing an usable software system with an interactive user interface(UI) so that doctors or patients can utilize it practically for detecting PCOS. Therefore, in addition to performing research on PCOS prediction, ML-based applications can be built in the future and used in real-world scenarios. Additionally, these software systems' usability and effectiveness can be assessed to see how well they enable physicians to make clinical decisions.

3.3.2.3 Emphasizing Deep Learning Unsupervised ML based techniques

The articles that have been reviewed in this SLR have all employed supervised machine learning techniques for PCOS detection. However, unsupervised ML techniques like clustering methods can be employed to reveal the patterns of the attributes behind PCOS as well as other useful information. Moreover, a limited numbers of articles have employed deep learning based methodologies which can be explored in a large scale in future. Deep learning architecture can also be employed to extract the significant features from the dataset for PCOS detection.

3.3.2.4 Enhancing Datasets and Country context for PCOS detection

The SLR findings reveal that, most of the articles have utlized datasets which are from the women of South-Asian subcontinent and specifically from India. But, PCOS is a global problem which is very common in women worldwide. A recent study conducted by Safiri et al. (Safiri et al. 2022) showed that, from 1990 to 2019, experts in 204 nations and territories discovered PCOS to be one of the most prevalent diseases among reproductive women. Therefore, the dataset that are to be employed in the future research in this domain should contain records from various country contexts throughout the world, so that the demographic and geographic impacts of this disease can be well investigated. Additionally, the ovarian USG imaging collection has to be enhanced further with good preservation in a clinical database so that researchers throughout the world may get more use out of it.

3.3.2.5 Investigating unknown reasons and attributes behind PCOS

All of the research so far in this field has been on the PCOS identification analyzing the patient records. However, the pathogenesis of PCOS is not entirely understood by the clinicians yet, which may be influenced by genetics, lifestyle, and a patient's lack of necessary micro-nutrients (Tefagh et al. 2022). Therefore, it's important to look into the causes of this anomaly as well as any other contributing factors for PCOS. Also, the clinicians found PCOS to be linked with various other metabolic and physiological disorders like obesity, insulin resistance (IR), ovarian cancer etc. (Condorelli et al. 2017). The researchers can focus on investigating the association between these anomalies and PCOS using machine learning based techniques (M. N. Islam, Mustafina, et al. 2022).

3.4 Chapter Summary

The existing relevant studies concentrating on the use of various computer-assisted methodologies for PCOS detection have been thoroughly examined in this systematic literature review to explore the state-of-art viewpoints, identify the shortcomings and determine the possible future study areas. The relevant articles have been selected meticulously using PRISMA methodology with distint inclusion-exclusion criteria. To address the first research objective (R01) the purpose, goals, and publication histories of the existing studies relevant to PCOS detection have been investigated. The data sources, type and attributes utilized in the examined studies with its relevance to research scopes have been investigated to address second research objective(RO2). The SLR also investigated into different computer-assisted methods, algorthims utilized in different studies and evaluation approaches that have been applied to automatically identify PCOS with an aim to fulfil third research objective (RO3). Furthermore, for addressing fourth research objective (RO4) the study has summarized the overall findings from the data synthesis from selected studies as well as the research gap. Finally, some suggestions have been provided concerning possible future study areas in the field of autonomous PCOS detection with enhanced efficiency with a goal to address fifth research objective (RO5).

One drawback of this literature review is that distinct exclusion-inclusion criteria may result in the exclusion of certain relevant and important publications; such as some publications could not be considered because the whole text was not available or they were written in a language other than English. Additionally, some associated articles might not be included because of the fixed set of lists of keywords used to search for articles. The authors of this study intend to undertake a more thorough review in the future, incorporating newer, more pertinent research papers as well as PCOS-related clinical studies.

Therefore, it is obvious that there are a vast array of potential areas for research in this domain, and the findings of this review will undoubtedly aid scholars in understanding the depth and breadth of existing studies, requirements and limitations to conduct more experiments and studies for PCOS detection in future. Additionally, this SLR also will aid the scholars and healthcare professionals to address the usefulness of different computer-assisted techniques, such as machine learning, image processing etc. in real-time decision-making for illness diagnosis at an early stage.

CHAPTER 4

PCOS DETECTION USING SYMPTOM DATA

The chapter discusses the PCOS detection procedure with multi-level stacked ensemble machine learning technique using patients' symptom data. The chapter firstly includes the materials and methods of the proposed technique which is subdivided into data acquisition, analysis and visualization; data preprocessing, feature selection and exploring machine learning models. After that, the results of the proposed methodology with the critical comparative analysis has been demonstrated; which includes performance comparison between the proposed technique with other ML techniques with varied numbers and set of features. Finally, the result findings has been discussed elaborately.

4.1 Phases of PCOS detection with USG image data

In this study, an extended ensemble machine learning classifier has been proposed, trained and tested using patient's most significant set of symptom data with an aim to differentiate between PCOS and non-PCOS patients. A framework of the research methodology is presented in figure 4.1. The research has been conducted through several phases to investigate and prioritize the optimum collection of features essential for predicting PCOS as well as to discover the best performing classification model for PCOS detection using those features. The classification models' inputs are a collection of patient's symptom attributes, and the outputs would be binary responses indicating whether a patient has PCOS or not. After retrieving the data from the repository, the dataset has been analyzed using various visualization approaches to gain a detailed understanding of it and then, it has been meticulously pre-processed to transform it into a clean and suitable dataset that can be used for

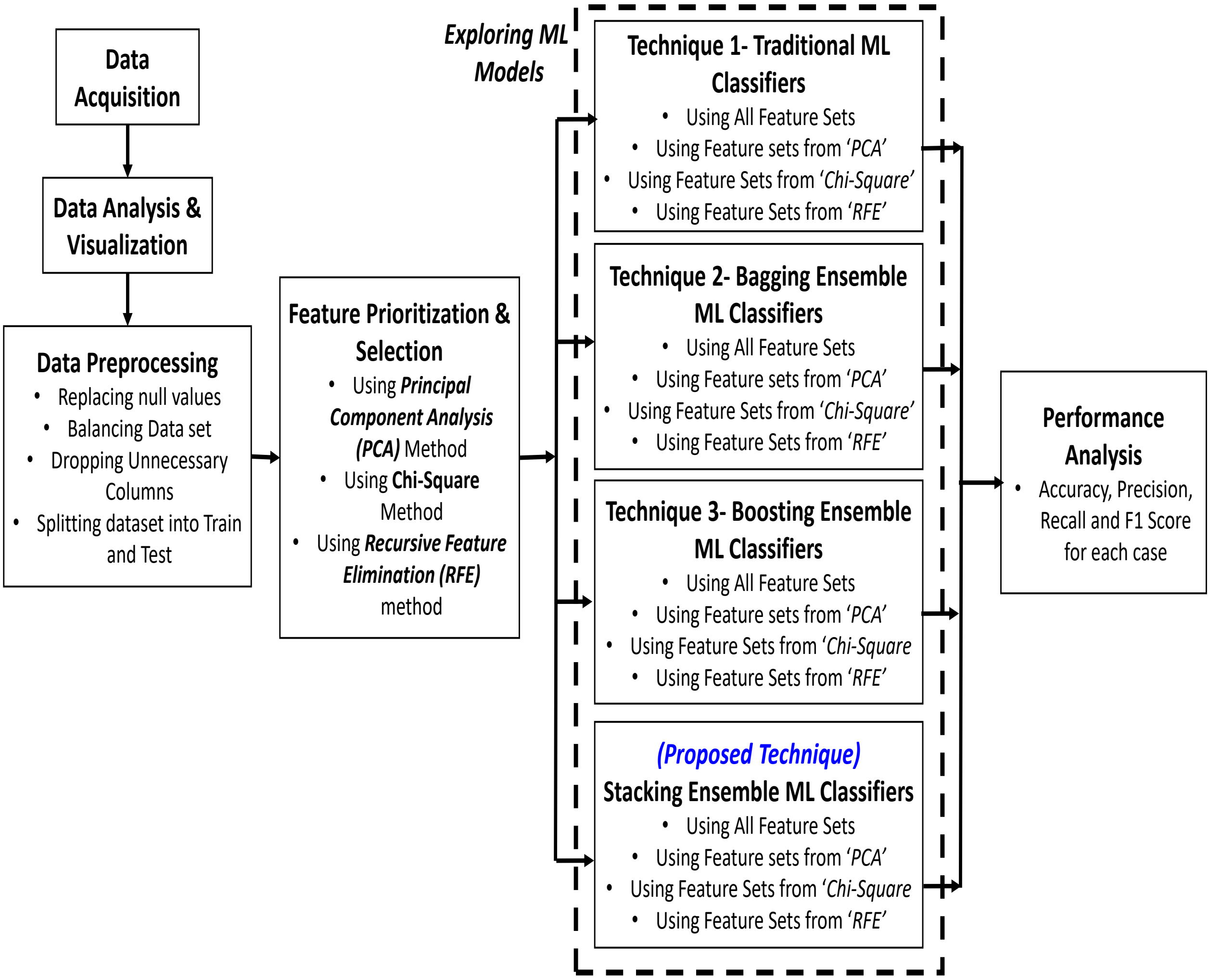


Fig. 4.1. Framework of research methodology

machine learning. Following that, several sets of reduced features with varied numbers of most significant attributes have been extracted using three different feature selection strategies. Then, for exploring different machine learning techniques several traditional as well as ensemble ML models have been trained and tested employing various sets of features and also a stacked ensemble model has been proposed. A comparative study of several classification models and feature prioritizing techniques has also been performed through different performance matrices to evaluate the efficacy of the classifiers. The methodological procedures that have been used in this research are detailed in the following subsections.

4.1.1 Data Acquisition, Analysis and Visualization

A dataset containing the symptoms along with the PCOS diagnosis findings of patients has been utilized here as the training data for supervised machine learning models, for which a publicly available data collection of PCOS patients from 'Kaggle' (Vedpathak 2020) has been selected. The dataset has been thoroughly investigated for better understanding before employing them for training purpose. The primary analysis of the PCOS records shows that,the dataset comprises a total of 541 records of female patient's data with 45 columns containing various types of clinical information related to PCOS anomaly. The feature list of the dataset have been shown in Table 4.1. One of the columns named 'PCOS(Y/N)' has a PCOS diagnosis outcome with 'Yes' and 'No' values indicating whether or not the patient has PCOS. This feature column has been considered as the target column for the training in this study. When the values of this column are counted, it has been observed that there is an uneven distribution of positive and negative outcomes, as there are 364 entries with 'No' indicating 'NO PCOS' and 177 entries with 'Yes' indicating 'PCOS'.

Table 4.1: All attributes of the dataset

Feature	Feature	Feature
Sl. No.	FSH(mIU/mL)	Hair loss(Y/N)
Patient File No.	LH(mIU/mL)	Pimples(Y/N)
Age (yrs)	FSH/LH	Fast food (Y/N)
Age (yrs)	Hip(inch)	Reg.Exercise(Y/N)
Height(Cm)	Waist(inch)	BP Systolic (mmHg)
BMI	Waist:Hip Ratio	BP Diastolic (mmHg)
Pulse rate(bpm)	TSH (mIU/L)	Follicle No. (L)
RR (breaths/min)	AMH(ng/mL)	Follicle No. (R)
Hb(g/dl)	PRL(ng/mL)	Avg. F size (L) (mm)
Cycle(R/I)	Vit D3 (ng/mL)	Avg. F size (R) (mm)
Cycle length(days)	PRG(ng/mL)	Endometrium (mm)
Marraige Status (Yrs)	RBS(mg/dl)	Blood Group
Pregnant(Y/N)	Weight gain(Y/N)	Skin darkening (Y/N)
No. of abortions	Body hair growth(Y/N)	Unnamed
I beta-HCG(mIU/mL)	II beta-HCG(mIU/mL)	PCOS (Y/N)

Again, the relationship between target column with other attributes has been examined using various visualization approaches. For example, the Figure 4.2(a) shows the age distribution of the patient records in the dataset which depicts that the records comprises information on

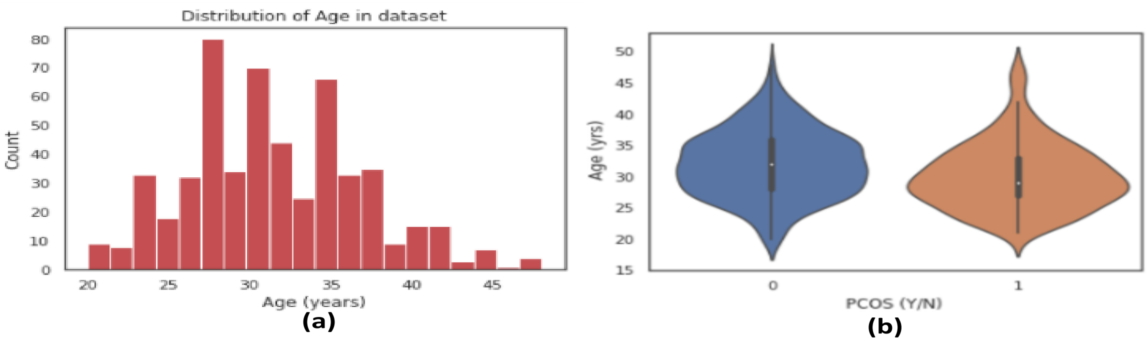


Fig. 4.2. (a) Distribution of Age in Dataset; (b) Relationship of 'Age' attribute with target column

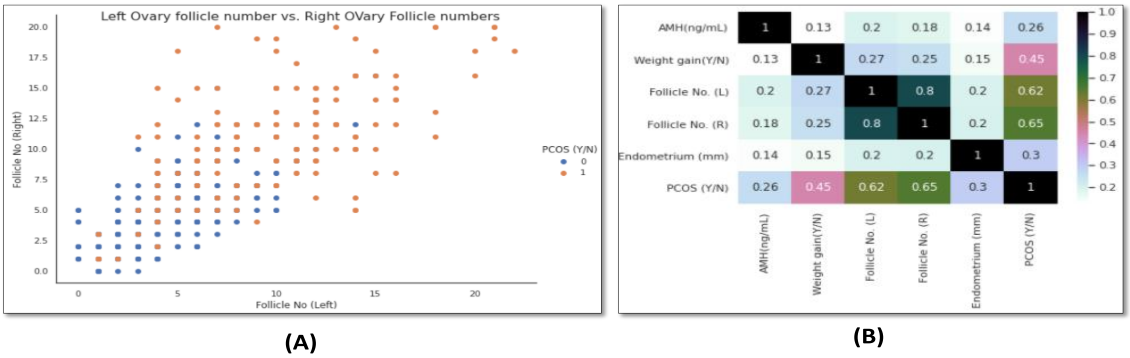


Fig. 4.3. (A) Scatter plot of Left ovary vs Right ovary Follicle numbers with respect to target attribute; (B) Correlation heatmap of some attributes

women aged 20 to 50 years old and the Figure 4.2(b) shows a violin plot of 'PCOS(Y/N) vs. Age(yrs)' which depicts the age range of women with and without PCOS in the dataset. Another example of data visualization has been illustrated in Figure 4.3 (a) where the number of follicles in the left ovary vs. right ovary in relation to the goal attribute 'PCOS(Y/N)' has been plotted, demonstrating that a larger number of follicles in both the left and right ovary yields the most positive PCOS outcomes. Furthermore, correlation study between the different attributes has been performed using a correlation heatmap to statistically analyze the strength of the relationship between the features, as an example illustrated in Figure 4.3 (b). The correlation value ranges from 0 to 1, where with a greater correlation value indicating that the features are highly correlated to each other.

4.1.2 Data Preprocessing

In this work, the dataset have been critically analyzed and preprocessed before using them into machine learning models to address the flaws and irregularities in the datasets such as missing or contradictory data samples, inconsistencies, noise, and other issues. The following steps have been employed for preprocessing the dataset:

Firstly, for data preprocessing the null values have been handled. Features with too many null or missing values have been completely removed from the dataset because they don't give any useful information; for example, the feature 'Unnamed' in the dataset contains 539 null values for which it has been eliminated. The features comprising a few null values have been substituted with other relevant values; for example, 'Marriage Status(Yrs)' and 'Fast Food (Y/N)' contains a few null values which have been replaced with 0.

Secondly, the data balancing has been done for making the classes equally distributed for training, as the dataset has imbalanced target attribute with 364 records of non-PCOS and 177 records of PCOS patients. Therefore, the dataset has been over-sampled using 'Synthetic Minority Oversampling Technique (SMOTE)' method that generates a synthetic sample of a minority class to eliminate the imbalance in the target attribute values (Rok and Lusa 2013). The mathematical formula followed for SMOTE method has been shown in equation (2), where x_{sample} is the sample generated from minority class value x and x_{random} is a randomly chosen value among the nearest neighbors of x with $0 \leq \eta \leq 1$. As a result of SMOTE, the dataset instances here have increased to 728 records, including 364 positive and negative PCOS diagnostic results.

$$x_{sample} = x + \eta(x_{random} - x) \tag{4.1}$$

The third step has been to Drop Unnecessary Columns. At this step, superfluous or duplicated columns have been removed in order to improve forecasting accuracy. One of two columns giving the same information has been kept, while the other has been deleted from the dataset. For example, 'I beta-HCG(mIU/mL)' numerical column and 'II beta-HCG(mIU/mL)' categorical attribute provides same information from which 'II beta-HCG(mIU/mL)' has been discarded. Also the unnecessary columns 'Sl. No.' and 'Patient File No.' have been discarded from the dataset as they contain simply the serial numbers, patient's file no which can be ignored for further analysis.

The next step is data normalization in which the values of the dataset are normalized using the MinMax Scalar approach to reduce the influence of variance in measurement units of different features and eliminate attribiute bias with sensitivities (*sklearn.preprocessing.MinMaxScaler* 2022). The MinMax scaller follows the equation (2) for rescaling the values of the feature range between 0 to 1. In the equation (2), x_{scaled} is the rescaled value generated from the original vale x where x_{min} and x_{max} are the minimum and maximum values of that attribute A_i.

$$x_{scaled} = (x - x_{min}) \div (x_{max} - x_{min}) \tag{4.2}$$

Finally, as the last step of data preprocessing, the dataset has been divided into train and test datasets for applying them to classification models of machine learning, with 30% of the instances randomly assigned to the test dataset and the remaining 70% assigned to the train dataset. The target feature has also been isolated from the other attributes, thus the 40 attributes are recorded as X_train or X_test arrays, and the target attribute 'PCOS(Y/N)' has been saved as y_train or y_test arrays.

4.1.3 Feature Selection

Feature selection is an efficient method for picking the most significant attributes and avoiding unimportant features to improve the prediction capacity and accuracy of machine learning algorithms (Maza and Touahria 2019). It is the process of exploring the best subset(s) of features to assure the finest potential data description. In this study, the dataset contains 40 attributes after preprocessing, which may lower the accuracy of the classifier if all of the less significant ones are taken into account. Thus, the features in this context have been prioritized and selected rigorously using three types of feature selection techniques to find out the optimal set of features from the PCOS data set. The techniques have been described hereafter:

- **Chi-Square Technique:** Chi-square feature selection technique is one of the most frequent and helpful feature selection strategies used in machine learning (Rustam and Ariantari 2018). It conducts a numerical test that calculates deviation from the anticipated distribution when the feature event is independent to the class value and prioritizes features by examining the relationship between them (Thaseen and C. A. Kumar 2017). The formula for the chi-sqaure feature selection has been shown in equation (3). Here, the prioritized features are chosen according to the best scores of χ^2. In case of implementation, the python 'SelectKBest' function has been utilized, which implemented the chi-square numeric test with $k = n$ and then picked n features from the dataset's 40 features based on the highest scores.

$$\chi^2 = \sum_{i=1}^{n} \frac{(ObservedValue_i - ExpectedValue_i)^2}{ExpectedValue_i} \tag{4.3}$$

- **Principal component analysis (PCA) Technique:** The second type of technique that has been used for feature selection in this study is the Principal component analysis (PCA) method, which is an efficient dimension reduction tool for feature prioritization utilizing numerical analysis which is accomplished by assessing the correlation between characteristics in order to determine the most important or principal components (Omuya, Okeyo, and Kimwele 2021; Banerjee, Gupta, and J. Saha 2018). PCA maps and reconstructs the original n-dimensional features to the required k-dimensional features $(k < n)$, where the k-dimensional features are new orthogonal attributes termed as principle components that minimize data redundancy to accomplish the dimension reduction goal (Zhao et al. 2019). In this scenario, the python 'PCA' function from Scikit-learn has been utilized with the PCA variance, to determine the most important n features.

- **Recursive Feature Elimination (RFE) Technique:** Recursive Feature Elimination, or RFE, is an efficient wrapper-type strategy that has been utilized in this study for removing features from a training sample for feature selection which ranks the set of attributes and eliminates them at the bottom that contribute the lowest to the categorization (Zeng, Y.-W. Chen, and Tao 2009). This approach is basically a recursive process that employs several machine learning techniques at its foundation, wrapped in the RFE methodology, and therefore feature importances are calculated at each iteration, with the least relevant one being eliminated to pick the prioritized features (Richhariya et al. 2020; Granitto et al. 2006). The RFE function from the RFE class provided by the scikit-learn Python machine learning library has been employed here for implementation

To explore the highly significant attributes that would yield the best performing accuracy when used in machine learning models, each type of feature selection approach selects the top 35, 30, 25, and 20 features from the PCOS dataset of 40 features. The algorithm followed for extracting the reduced set of features from the dataset in this study is shown in Algorithm 1. Then employing those different sets of features the machine learning classifiers are trained, tested and evaluated through different performance metrics.

Algorithm 1 Pseudo Code for Feature Selection

Input: AF: Set of initial all features;
$numF$: Number of reduced features wanted
$FStech$: Feature selection technique (Chi^2, PCA, RFE)
Output: RF_{numF}: Reduced feature set with '$numF$' number of features [1] **forall** f_i in AF **do**

end
relative importance of f_i $\Leftarrow$ $FStech\ relevanceValues[f_i]$ $\Leftarrow$ relative importance of f_i
$AF \Leftarrow sort(relevanceValues[f_i])$ $RF_{numF} \Leftarrow AF.take(numF)$ **return** RF_{numF}

4.1.4 Exploring Machine Learning Models

Classification is a machine learning technique that uses a model learned from training data to forecast the category of samples and therefore maps or classifies data instances into the associated class labels which have been predefined in the provided dataset (Han, Pei, and Kamber 2011; M. N. Islam and A. N. Islam 2020). In this study, for training the machine learning models with an aim to differentiate between PCOS and non-PCOS classes from their symptom data, four types of techniques have been employed (see Figure 4.1). The predictive models have been trained, tested and evaluated using different sets of features from the dataset. The machine learning techniques have been discussed briefly below:

Table 4.2: A summary of traditional machine learning classifiers used for different healthcare predictive studies

Classifiers	Brief Description	Examples of healthcare predictions	References
Logistic Regression	A probabilistic-based statistical model in which the classifier assesses the association between the dependent variable as target class and independent variables or features for a given dataset using a logistic function (LaValley 2008)	Chronic disease prediction, ovarian cancer classification, Alzheimer's disease detection etc.	Nusinovici et al. 2020, Octaviani, Rustam, and Siswantining 2019, Xiao et al. 2021
Support Vector Machine	A hyperplane is chosen, which is a line that can discover the coefficients, separate samples in the variable space with the best detachment of the classes Keerthi et al. 2001	PCOS detection, heart disease diagnosis, cervical cancer detection etc.	Sengur 2012, Bharati, Podder, and M. R. H. Mondal 2020, J. Zhang and Liu 2004
Decision Tree	Estimates entropy and information gain for each attribute over a provided training sample and analyzes each feature at each node of a top-down tree for classification (Quinlan 1986)	Parkinson's disease identification, COVID-19 diagnosis, coronary artery disease diagnosis etc.	Syapariyah, Saifudin, Desyani, et al. 2020, Yoo et al. 2020, Ghiasi, Zendehboudi, and Mohsenipour 2020
K-Nearest Neighbour	It's a instance-based learning that considers local approximation presuming that similar data are close together & computation is conducted until classification (Sarker, Kayes, and Watters 2019)	Diabetes detection, chronic kidney disease prediction, Ovarian cancer classification etc.	Suyanto et al. 2022, Devika, Avilala, and Subramaniyaswamy 2019, Alqudah 2019
Naive Bayes	A fundamental probabilistic based classification strategy for predicting class membership probability by computing the likelihood of membership for each category (M. J. Islam et al. 2007)	Breast cancer detection, brain tumor detection, thyroid detection etc.	Kharya and Soni 2016, Zaw, Maneerat, and Win 2019, Chandel et al. 2016

4.1.4.1 Existing Machine Learning Techniques

- Traditional ML Classifiers: Numerous classification methods with the potential to make forecasts have indeed been developed in the predictive analytics literature, but some classical machine learning classification techniques have been commonly used in various studies to estimate various clinical anomalies M. N. Islam and A. N. Islam 2020. Here, technique 1 employs five kinds of well-known and widely utilized traditional machine learning classification techniques with fundamental algorithmic structure which are appropriate to this target area. The models are Logistic Regression classifier, Support Vector Machine classifier, Decision Tree classifier, K-Nearest Neighbour classifier and Naive Bayes classifier. These machine learning classifiers have been applied extensively in a variety of healthcare-related predictive studies. Table 4.2 shows a summary of these traditional machine learning models used in various clinical prediction related studies.

Table 4.3: A summary of bagging and boosting ensemble machine learning classifiers used for different healthcare predictive studies

Type	ML Classifiers	Brief Description	Examples of healthcare predictions	References
Bagging ensemble	Random Forest classifier	Integrates bootstrap aggregation (bagging) and random feature selection to create a set of decision trees with controlled variation that can anticipate the corresponding output activity class (Breiman 2001)	PCOS detection, lymph disease diagnosis, thyroid disorder anlysis etc.	Tiwari et al. 2022, Azar et al. 2014, Mishra et al. 2021
Boosting ensemble	Gradient Boosting classifier	It is an ensemble forward learning model which eliminates all weaker predictors in favor of a stronger one using an upgraded version of the decision tree, in which each successor is selected using the refined structure score, gain computation, and advanced approximations (Shrivastav and Jha 2021)	Lung cancer detection, diabetes diagnosis, Leukemia prediction etc.	Chandrasekar et al. 2022, Bahad and Saxena 2020, Deif, Hammam, and Solyman 2021
	eXtreme Gradient (XG) Boosting	This approach is scalable and efficient form of gradient boosting that improves on two fronts: tree construction speed and a novel distributed algorithm for tree searches (T. Chen and Guestrin 2016)	Heart disease detection, chronic kidney disease diagnosis, breast cancer detection etc.	Ashish, S. Kumar, and Yeligeti 2021, Ogunleye and Q.-G. Wang 2019, M. S. K. Inan, R. Hasan, and Alam 2021
	Adaptive Boosting classifier	It's an adaptive classifier that leverages the results of various weak learning algorithms to substantially enhance performance and provide an effective predictor for the boosted classifier's final output (Freund, Schapire, et al. 1996)	Endometrial cancer prediction, Hepatitis disease detection, cancer classification etc.	X. Wang and R. Zhang 2022, Akbar et al. 2020, Lu et al. 2019
	Categorical Gradient (CAT) Boosting	It is an implementation of Gradient Boost classifier that employs ordered boosting with categorical features and uses binary decision trees as underlying predictors (Prokhorenkova et al. 2018)	Parkinson's disease prediction, COVID-19 detection from blood samples, diabetes risk prediction etc.	Al-Sarem et al. 2021, Abayomi-Alli et al. 2022, P. S. Kumar et al. 2021

- Bagging Ensemble ML classifiers:

  A bagging classifier or bootstrap aggregation classifier is an ensemble method that fits multiple base classification models on randomized subsets of the dataset with the same weights given to each models and then aggregates their individual predictions to generate a final result (Yaman and Subasi 2019). In this study, Random Forest classifier has been used for classification as a type of bagging classifier which is created based on the aggregation of numerous decision tree base classifiers. During the evolution of a decision tree, Random Forest employs random subset or feature projection which means rather than using all of the parameters in one tree, each decision tree in Random Forest selects only a subset of variables at every prospective splits (Lee, Ullah, and Ran Wang 2020). A number of researchers have used random forest classifier successfully to the various domain of healthcare predictive analysis. A brief summary in this regard has been shown in Table 4.3.

- Boosting Ensemble ML classifiers:

  Boosting is an ensemble machine learning approach in which a random sample data is chosen, fitted with a model, and then trained in a sequential manner, combining a set of weak learners into a strong learner with an aim to minimize training errors, with every model attempting to compensate for the shortcomings of the previous model (W. Chen et al. 2021). Based on the different ways of producing and aggregating weak learners during the sequential approach, boosting algorithms can be categorized into different types. In this study, four types of widely utilized variations of boosting ensemble technique (N. I. Khan et al. 2020) have been employed which are : Gradient(Grad) Boosting classifier, Adaptive(Ada) Boosting Classifer, eXtreme Gradient(XG) Boosting classifier and CAT Boosting classifier. These classifiers have been considered here because they have been successfully applied to a range of challenges in the field of healthcare predictive modeling, as a brief summary shown in Table 4.3.

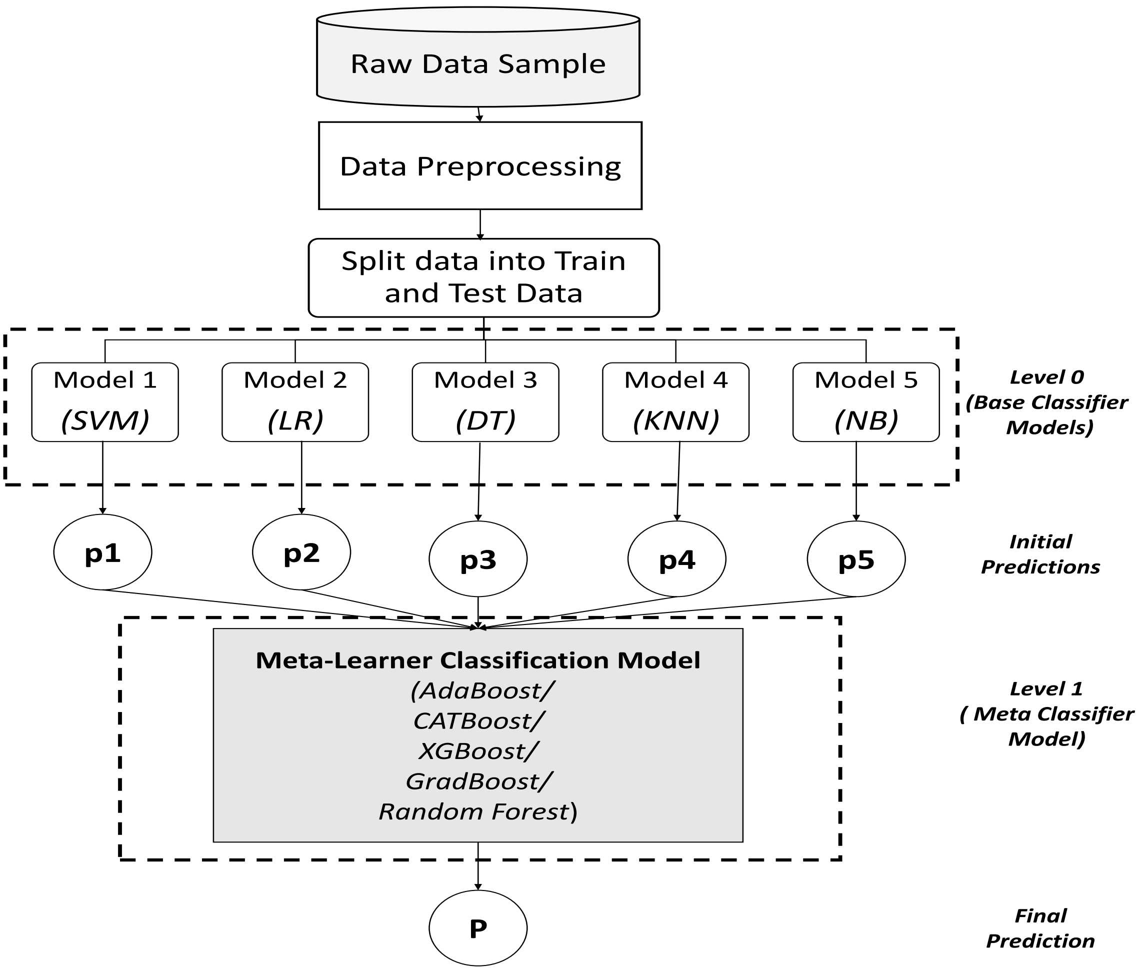


Fig. 4.4. Basic Framework of the Proposed Stacking Ensemble Technique

4.1.4.2 Proposed Machine Learning classifiers

A stacking ensemble based ML classification approach has been proposed for predicting the PCOS or non-PCOS criteria in this study that differs from bagging and boosting approaches in the following perspectives : (a) it analyzes heterogeneous weak classifiers as well as learns them in parallel; (b) then aggregates them by training a meta-learner to produce a prediction based on the forecasts of the individual weak learners; and (c) thus, it minimizes variance and also enhances predictive force of the learning process (Zounemat-Kermani et al. 2021). The basic framework of the proposed stacking ensemble machine learning technique has been illustrated in Figure 4.4.

The proposed model is a multi-level stacked ensemble model where after preprocessing the raw data sample, it is being divided into train and test data and then initially sent to the base learners of level 0. At this phase, the five types of widely utilized traditional machine learning classifiers have been considered to be the weak learners or base classifiers at level 0 of the stacked model, which are: Logistic Regression(LR), Support Vector Machine(SVM), Decision Tree(DT), K-Nearest Neighbour(KNN) and Gaussian Naive Bayes(NB) classifiers (see Table 4.2). Each of these base models gets trained independently using their respective prediction algorithms, resulting in forecasts symbolized p1, p2, p3, p4, and p5 in Figure 4.4. Following that, the predictions acquired from the level 0 models are fed into level 1 where a single classification model, or meta-learner, learns to generate final prediction from it. The meta-learner has been created at level 1 employing one stronger machine learning classifier. Here, while keeping the same base models at level 0, five types of classifiers have been explored as meta learners in level 1 which resulted in five varieties of the proposed architecture with an aim to explore the best performing one. The meta learner is one of the five types of bagging or boosting classifiers outlined before in Table 4.3 (Random Forest, Adaptive Boosting, Gradient Boosting, CAT Boosting and XGBoosting) that is ultimately trained on top of level 0 to generate the final output based on the predictions returned by the base models. Thus, an extended stacked ensemble ML classifier has been proposed incorporating five types of traditional classifiers as base models and one boosting or bagging type of classifier as meta learner with an aim to differentiate between PCOS and non PCOS patients.

4.2 Results and Findings

In this study, to evaluate and compare the efficacy of the predictive models for PCOS detection, total four types of ML techniques have been performed employing fifteen varieties of ML classifiers including the traditional (five models), bagging (one model), boosting(four models) and proposed (five models) techniques. All the experiments have been simulated using patient symptom dataset for classifying the records into PCOS and non-PCOS criteria. Furthermore, to explore the optimum and most significant attributes from the dataset, three types of feature selection methods(Chi-square, PCA and RFE) have been employed which picked the top 35,30,25 and 20 features out of the 40 features of the dataset. Each ML model's performance have then been evaluated employing these different sets of features acquired from feature selection techniques.

The performance of different varieties of machine learning algorithms utilizing different sorts of feature sets is analyzed using four performance measures, which are Accuracy, Precision, Sensitivity (recall), and F1 score, to investigate the efficacy of the prediction analysis (S. Xu et al. 2022; Suha, Akhtaruzzaman, and Sanam 2022). The performance metrics are primarily based on a comparison of anticipated and actual values that investigates number of correct and incorrect predictions from the training sample, which is divided into four categories: True Positive (TP) that is both the true and predicted values are positive; True Negative(TN) in which both the original and the anticipated values are negative.; False Positive(FP) a where the actual value is negative but the anticipated result is positive and lastly False Negative(FN) where the actual value is positive, but the predicted result is negative . Based on these evaluations, the performance measures utilized here can be stated as equations (4), (5), (6), and (7):

$$Accuracy = (TP\ +\ TN) \div (TP + TN + FP + FN) \tag{4.4}$$

$$Precision = TP \div (TP + FP) \tag{4.5}$$

$$Recall(Sensitivity) = TP \div (TP + FN) \tag{4.6}$$

$$F1 - score = 2 * (Precision * Sensitivity) \div (Precision + Sensitivity) \tag{4.7}$$

The findings of this rigorous evaluation process have been shown in Table 4.4, 4.5, 4.6, 4.7; where Table 4.4 shows the accuracy, Table 4.5 shows the precision, Table 4.6 shows the recall and Table 4.7 shows the F1-score of different models using different sets of features.

Table 4.4: Accuracy Comparison of ML models using different set of features

Type	Classification Models	**40 Feat.**	**35 Feat. Accuracy**			**30 Feat. Accuracy**			**25 Feat. Accuracy**			**20 Feat. Accuracy**		
			Chi^2	PCA	RFE	Chi^2	PCA	RFE	Chi^2	PCA	RFE	Chi^2	PCA	RFE
Trad Tech	SVM	0.507	0.89	0.511	0.534	0.871	0.911	0.893	0.708	0.921	0.679	0.609	0.583	0.611
	Log. Reg	0.872	0.89	0.886	0.895	0.92	0.911	0.893	0.702	0.921	0.688	0.714	0.611	0.711
	DecisionTree	0.836	0.822	0.868	0.863	0.804	0.78	0.879	0.716	0.864	0.713	0.653	0.664	0.65
	KNN	0.685	0.816	0.667	0.685	0.816	0.864	0.813	0.622	0.879	0.736	0.696	0.716	0.707
	NaiveBayes	0.868	0.84	0.863	0.863	0.859	0.766	0.864	0.74	0.808	0.76	0.565	0.585	0.564
Bagging Tech	RandomForest	0.889	0.89	0.906	0.902	0.89	0.907	0.902	0.851	0.916	0.86	0.708	0.793	0.8
Boosting Tech.	GradBoosting	0.872	0.877	0.883	0.893	0.89	0.888	0.93	0.853	0.874	0.832	0.777	0.715	0.706
	XG Boosting	0.89	0.877	0.897	0.864	0.881	0.907	0.897	0.853	0.85	0.864	0.689	0.675	0.715
	AdaBoosting	0.886	0.896	0.869	0.883	0.853	0.902	0.902	0.871	0.864	0.869	0.756	0.8	0.686
	CATBoosting	0.9	0.863	0.841	0.832	0.9	0.916	0.916	0.865	0.879	0.85	0.789	0.799	0.725
Proposed Tech (stacking)	Meta learner-Grad Boost	0.927	0.922	0.932	0.918	0.926	0.953	0.943	0.883	0.957	0.911	0.853	0.832	0.86
	Meta learner-XGBoost	0.913	0.918	0.918	0.922	0.926	0.935	0.924	0.859	0.942	0.893	0.871	0.822	0.897
	Meta learner-AdaBoost	0.922	0.934	0.927	0.936	0.946	0.931	0.938	0.883	0.942	0.907	0.808	0.802	0.893
	Meta learner-CATBoost	0.913	0.89	0.932	0.918	0.933	0.943	0.933	0.908	0.947	0.893	0.802	0.825	0.802
	Meta learner-RandForest	0.909	0.883	0.913	0.927	0.924	0.918	0.925	0.89	0.925	0.916	0.802	0.812	0.807

Table 4.5: Precision Comparison of ML models using different set of features

Type	**Classification Models**	**40 Feat.**	**35 Feat. Precision**			**30 Feat. Precision**			**25 Feat. Precision**			**20 Feat. Precision**		
			Chi^2	PCA	RFE	Chi^2	PCA	RFE	Chi^2	PCA	RFE	Chi^2	PCA	RFE
Trad Tech	SVM	0.65	0.874	0.514	0.541	0.855	0.912	0.894	0.799	0.921	0.68	0.61	0.582	0.613
	Log. Reg	0.872	0.878	0.887	0.897	0.91	0.912	0.893	0.791	0.921	0.69	0.715	0.621	0.711
	DecisionTree	0.837	0.802	0.868	0.863	0.783	0.785	0.879	0.715	0.865	0.715	0.654	0.66	0.651
	KNN	0.697	0.798	0.667	0.685	0.795	0.865	0.816	0.602	0.882	0.736	0.698	0.715	0.71
	NaiveBayes	0.874	0.841	0.869	0.865	0.873	0.805	0.865	0.741	0.829	0.76	0.565	0.586	0.565
Bagging Tech	RandomForest	0.888	0.878	0.907	0.902	0.878	0.912	0.902	0.85	0.917	0.861	0.708	0.792	0.802
Boosting Tech	GradBoosting	0.873	0.866	0.885	0.894	0.88	0.888	0.93	0.847	0.876	0.836	0.775	0.716	0.706
	XG Boosting	0.891	0.879	0.898	0.865	0.882	0.909	0.897	0.886	0.852	0.865	0.688	0.67	0.715
	AdaBoosting	0.886	0.886	0.87	0.883	0.837	0.902	0.902	0.864	0.866	0.869	0.755	0.822	0.67
	CATBoosting	0.901	0.872	0.843	0.836	0.89	0.92	0.917	0.85	0.88	0.854	0.78	0.796	0.724
Proposed Tech (stacking)	Meta learner Grad Boost	0.927	0.921	0.931	0.918	0.925	0.956	0.945	0.868	0.952	0.912	0.85	0.835	0.862
	Meta learner XGBoost	0.914	0.895	0.918	0.923	0.922	0.934	0.925	0.853	0.942	0.893	0.867	0.823	0.895
	Meta learner AdaBoost	0.922	0.925	0.927	0.936	0.948	0.942	0.937	0.872	0.931	0.908	0.81	0.802	0.893
	Meta learner CATBoost	0.914	0.864	0.932	0.918	0.935	0.944	0.933	0.897	0.947	0.893	0.801	0.826	0.806
	Meta learner RandForest	0.909	0.868	0.914	0.927	0.925	0.919	0.924	0.876	0.926	0.916	0.802	0.811	0.807

Table 4.6: Recall Comparison of ML models using different set of features

Type	Classification Models	40 Feat.	35 Feat. Recall			30 Feat. Recall			25 Feat. Recall			20 Feat. Recall		
			Chi^2	PCA	RFE	Chi^2	PCA	RFE	Chi^2	PCA	RFE	Chi^2	PCA	RFE
Trad Tech	SVM	0.513	0.886	0.506	0.531	0.864	0.911	0.893	0.79	0.922	0.67	0.612	0.583	0.622
	Log. Reg	0.872	0.878	0.886	0.896	0.914	0.911	0.893	0.79	0.921	0.68	0.725	0.62	0.71
	DecisionTree	0.835	0.805	0.867	0.863	0.778	0.78	0.879	0.8	0.867	0.73	0.655	0.66	0.655
	KNN	0.687	0.822	0.667	0.685	0.809	0.864	0.813	0.61	0.887	0.74	0.698	0.725	0.711
	NaiveBayes	0.868	0.798	0.864	0.863	0.812	0.766	0.864	0.78	0.815	0.86	0.56	0.596	0.565
Bagging Tech	RandomForest	0.888	0.878	0.906	0.902	0.878	0.907	0.902	0.85	0.922	0.86	0.709	0.793	0.803
Boosting Tech	GradBoosting	0.873	0.86	0.883	0.893	0.873	0.888	0.93	0.82	0.871	0.83	0.777	0.717	0.71
	XG Boosting	0.89	0.878	0.897	0.864	0.881	0.907	0.897	0.79	0.853	0.86	0.689	0.673	0.712
	AdaBoosting	0.889	0.882	0.869	0.883	0.837	0.902	0.902	0.85	0.862	0.88	0.754	0.823	0.671
	CATBoosting	0.9	0.869	0.841	0.832	0.877	0.916	0.916	0.85	0.886	0.85	0.781	0.794	0.724
Proposed Tech (Stacking)	Meta learner Grad Boost	0.927	0.92	0.932	0.918	0.929	0.956	0.943	0.87	0.952	0.91	0.851	0.822	0.86
	Meta learner XGBoost	0.914	0.904	0.918	0.923	0.921	0.934	0.924	0.83	0.942	0.89	0.877	0.826	0.894
	Meta learner AdaBoost	0.922	0.925	0.927	0.936	0.948	0.942	0.938	0.87	0.934	0.91	0.812	0.805	0.893
	Meta learner CATBoost	0.914	0.864	0.932	0.918	0.934	0.944	0.933	0.9	0.944	0.89	0.804	0.827	0.807
	Meta learner Rand Forest	0.909	0.877	0.914	0.927	0.925	0.919	0.923	0.88	0.922	0.92	0.802	0.815	0.81

Table 4.7: F1-Score Comparison of ML models using different set of features

Type	**Classification Models**	**40 Feat.**	**35 Feat. F1-Score**			**30 Feat. F1-Score**			**25 Feat. F1-Score**			**20 Feat. F1-Score**		
			Chi^2	PCA	RFE	Chi^2	PCA	RFE	Chi^2	PCA	RFE	Chi^2	PCA	RFE
Trad Tech	SVM	0.367	0.88	0.427	0.5	0.864	0.911	0.893	0.79	0.92	0.69	0.691	0.674	0.683
	Log Reg	0.872	0.878	0.886	0.895	0.914	0.911	0.893	0.79	0.92	0.68	0.712	0.678	0.711
	DecisionTree	0.835	0.804	0.868	0.863	0.778	0.78	0.879	0.79	0.86	0.71	0.623	0.62	0.664
	KNN	0.681	0.805	0.667	0.685	0.809	0.864	0.813	0.61	0.87	0.74	0.684	0.786	0.716
	Naive Bayes	0.867	0.812	0.863	0.863	0.812	0.766	0.864	0.71	0.81	0.86	0.543	0.555	0.585
Bagging Tech	RandomForest	0.889	0.878	0.91	0.902	0.878	0.907	0.902	0.85	0.91	0.86	0.781	0.685	0.853
Boosting Tech	GradBoosting	0.872	0.863	0.883	0.892	0.873	0.888	0.93	0.83	0.87	0.81	0.793	0.769	0.731
	XG Boosting	0.89	0.878	0.897	0.864	0.881	0.907	0.897	0.82	0.85	0.86	0.683	0.641	0.72
	AdaBoosting	0.887	0.0884	0.869	0.883	0.837	0.902	0.902	0.85	0.86	0.87	0.783	0.864	0.602
	CATBoosting	0.9	0.87	0.841	0.831	0.877	0.916	0.916	0.85	0.87	0.85	0.785	0.769	0.725
Proposed Tech (Stacking)	Meta learner Grad Boost	0.927	0.92	0.932	0.918	0.929	0.956	0.943	0.87	0.95	0.91	0.891	0.885	0.82
	Meta learner XGBoost	0.913	0.899	0.918	0.922	0.921	0.934	0.924	0.84	0.92	0.89	0.882	0.869	0.869
	Meta learner AdaBoost	0.922	0.925	0.927	0.936	0.948	0.942	0.938	0.87	0.93	0.91	0.814	0.801	0.888
	Meta learner CATBoost	0.913	0.864	0.932	0.918	0.934	0.944	0.933	0.89	0.94	0.89	0.809	0.824	0.81
	Meta learner Random Forest	0.909	0.872	0.913	0.927	0.925	0.919	0.923	0.87	0.92	0.92	0.811	0.829	0.818

4.2.1 Comparative performance analysis

Analyzing the evaluation results from Table 4.4, 4.5, 4.6, 4.7, it can be observed that, the performances of the classifiers enhance significantly using the proposed stacked ensemble techniques. For example, incorporating all the 40 features the best performance has been achieved using the proposed stacked ensemble classifier with Gradient boosting model as meta learner attaining 92.7% accuracy, 92.7% precision, 92.7% recall and 92.7% F1 score. Also it is noticiable that, each of the stacked ensemble models has acquired accuracy performance over 90% using the proposed technique with all features whereas the other models typically have less than or equal to 90% accuracy. Similar findings are also observed in case of using all the reduced set of features (set of 35,30,25 and 20 features) acquired from feature selection techniques, where the five varieties of proposed ML models outperform the other types of models in terms of all the performance metrics.

Figure 4.5 graphically illustrates the comparative analysis of the accuracy of different ML models incorporating different feature sets where Figure 4.5 (A) shows the accuracy of the models with features selected using chi-square method, Figure 4.5 (B) shows accuracy with PCA features and Figure 4.5 (C) shows comparative accuracy with RFE features. Each of the graphical representation compares the accuracy performances of four techniques with different classification models employed in this study using 40 features, 35 features, 30 features ,25 features and 20 features selected using chi-square (see 4.5 (A)), PCA (see 4.5 (B)) and RFE (see 4.5 (C)) feature selection method. From the graphical representation it is clearly visible that the performances of the proposed stacking ensemble models are comparatively higher than the other models in case of all types of feature selection methods. Thus, the results acquired from evaluating the ML techniques with different performance metrics clearly indicates that, the proposed stacking ensemble techniques provides a better performance for classifying the dataset into PCOS and non-PCOS classes.

4.2.2 Result Analysis of Feature Selection

The different feature selection methods utilized here have selected different sets of attributes employing their own methodologies. An example has been given in Table 4.8 which shows

Table 4.8: Top 25 dominant features prioritized by three types of feature selection methods

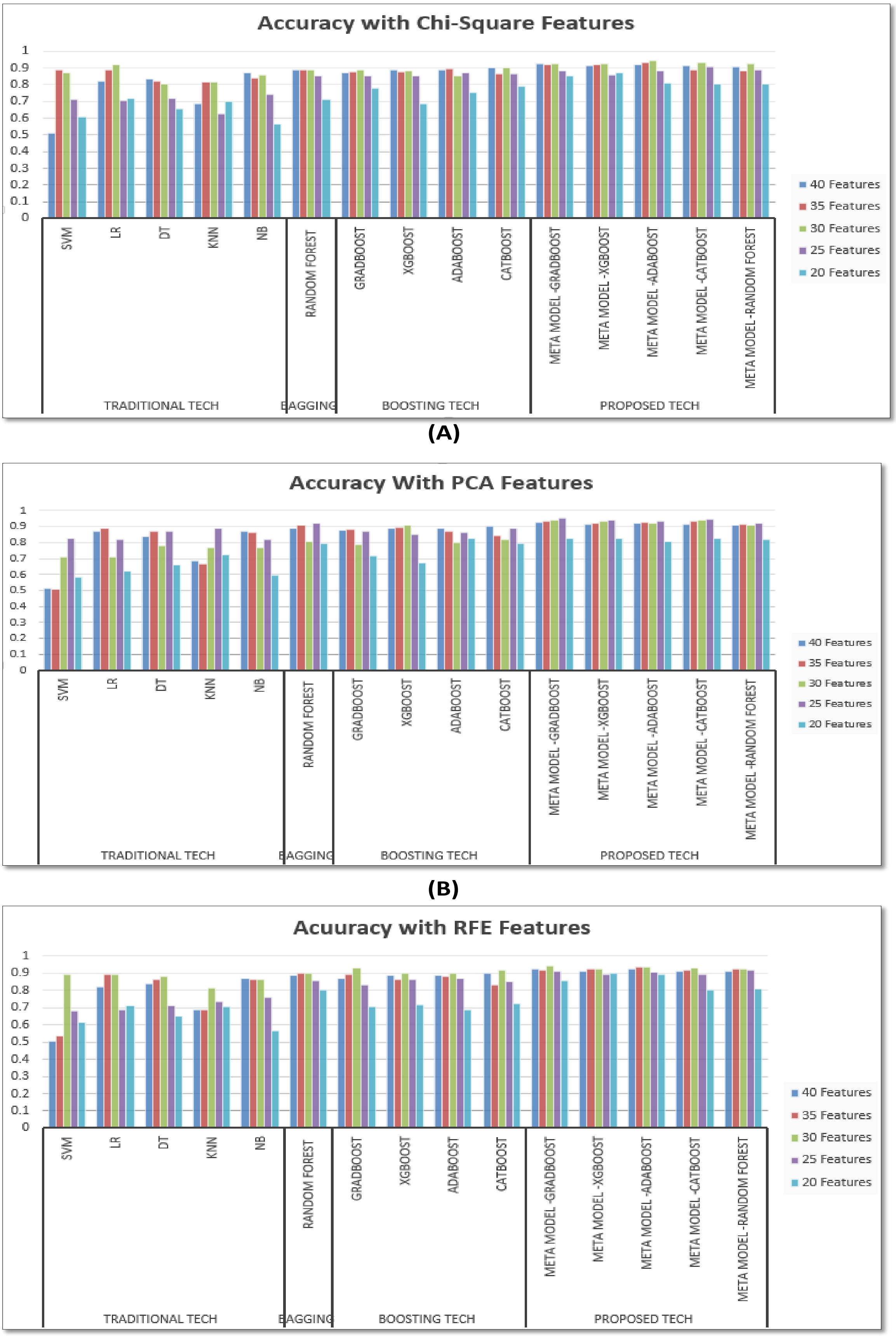
	Chi-Square	**PCA**	**RFE**
1	Age (yrs)	Weight (Kg)	Weight (Kg)
2	Weight (Kg)	BMI	Height(Cm)
3	BMI	WeightGainY/N	BMI
4	Cycle(R/I)	Waist(inch)	Marraige Sta(yr)
5	Cycle length	Hip(inch)	Cycle(R/I)
6	Marraige Sta.(yr)	hair growth-Y/N	Endometrium
7	Pregnant(Y/N)	Follicle No. (L)	Pregnant(Y/N)
8	No. of abortions	Fast food (Y/N)	Pulse rate(bpm)
9	LH(mIU/mL)	Skin dark(Y/N)	FSH(mIU/mL)
10	FSH(mIU/mL)	Follicle No. (R)	LH(mIU/mL)
11	Hip(inch)	Avg. F size (L)	TSH(mIU/L)
12	Waist(inch)	Avg. F size (R)	PRG(ng/mL)
13	AMH(ng/mL)	Cycle(R/I)	No. of abortions
14	Vit D3 (ng/mL)	Pimples(Y/N)	WeightGain
15	PRG(ng/mL)	Hair loss(Y/N)	hair growth-Y/N
16	WeightGain-Y/N	Height(Cm)	Skin dark(Y/N)
17	hair growth-Y/N	AMH(ng/mL)	Hair loss(Y/N)
18	Skin dark(Y/N)	Endometrium	Pimples(Y/N)
19	Hair loss(Y/N)	FSH/LH	Fast food (Y/N)
20	Pimples(Y/N)	Cycle length	Follicle No. (R)
21	Fast food (Y/N)	Hb(g/dl)	Follicle No. (L)
22	Reg.Exer.-Y/N	Vit D3 (ng/mL)	Cycle length
23	Follicle No. (L)	RBS(mg/dl)	Avg. F size (L)
24	Follicle No. (R)	Age (yrs)	Reg.Exer.-Y/N
25	Avg. F size (L)	BP Systolic	RR(breaths/min)

Table 4.9: Categorization of top 25 dominant features based on PCA technique

Feature Categories	Features	
Demographics	Age (yrs)	
Vital Signs	Weight Gain	Cycle(R/I)
	Body hair growth	Pimples
	Skin darkening	Hair loss
Patient History	BMI	Waist(inch)
	Cycle length	Hip(inch)
	Weight (Kg)	Height(Cm)
	BP Systolic	
Laboratory Diagnosis Outcomes	Follicle No(L)	Follicle No(R)
	Avg. F size (L)	Avg. F size (R)
	Endometrium Thickness	
Comorbidities	Hb(g/dl)	AMH(ng/mL)
	FSH/LH	RBS(mg/dl)
	Vit D3 (ng/mL)	

the top 25 features that have been picked using three types of feature selection methods (Md Mahbubur Rahman et al. 2022a). It is apparent from the table that the most essential attribute of the three techniques are evidently nonidentical. From the table it is observable that, the three set of top 25 features differs from each other such as both PCA and RFE has considered 'Endrometrium' (endrometrium thickness of follicles) as a significant feature but Chi-square technique has not selected it; on the other hand chi-square technique has selected 'Marriage Status(Yrs)' as an important features but PCA technique has not prioritized it; and so on. This results indicate that different feature selection techniques picks different combinations of features from the dataset and thus it is necessary to investigate which set of feature provides the best performance.

From the comparative evaluation with graphical representation in Figure 4.5, another significant finding is that, the accuracy of the models employing Chi-square and RFE feature selection methods gradually enhances when the number of features have been reduced from 40 features to 30 selected features; but then the performances start decreasing for the selected 25 and 20 features for almost all the models. The highest accuracy for most of the models employing chi-square and RFE feature selection method has been acquired with top 30 selected features. Here, the highest accuracy with Chi-square featue selection method has been achieved using stacking ensemble classifier with 'AdaBoost' model as meta learner



(C)

Fig. 4.5. Comparative accuracy analysis of ML models with different sets of (A)Chi-Square, (B)PCA and (C)RFE features

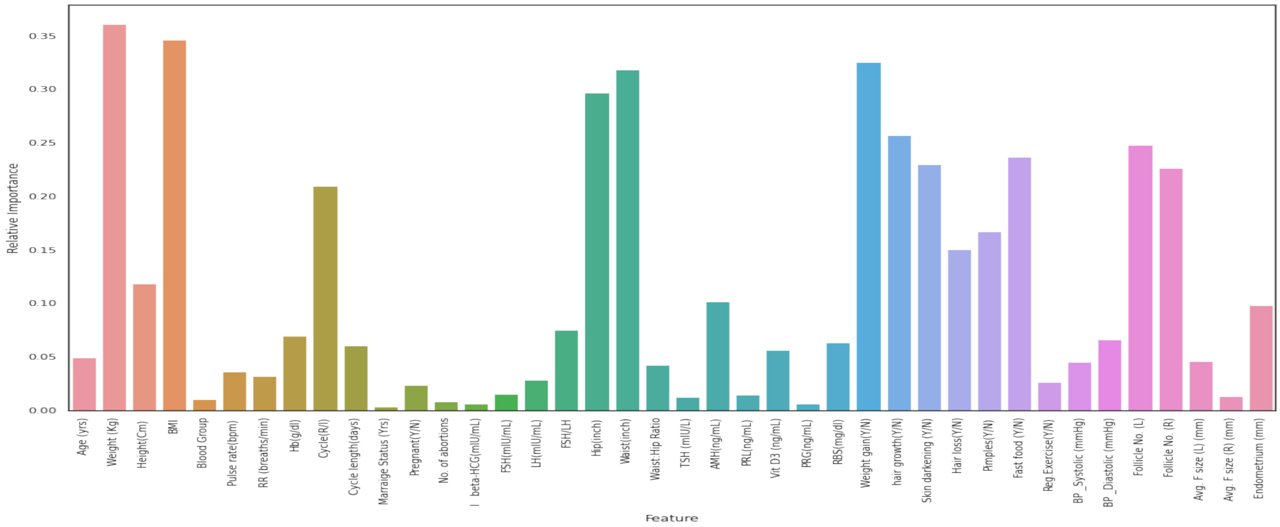


Fig. 4.6. Relative importance of features based on PCA technique

which is 94.6% using top 30 features; and the highest accuracy with RFE featue selection method has been achieved using stacking ensemble classifier with ‘GradBoost’ model as meta learner which is 94.3%.

However, when using the PCA feature selection approach, most of the models’ accuracy consistently improves with reduced features and has reached its peak with the top 25 features. Figure 4.6 graphically displays the relative importance of all the features of the dataset based on PCA technique. Using the top 25 features selected via the PCA approach, the maximum accuracy being 95.7% has been achieved in this context with a stacking ensemble classifier with the ‘GradBoost’ model as the meta learner. The most significant 25 attributes providing the best performance that has been explored using PCA technique is shown in Table 4.9. In this table the top selected features are further grouped based on the real-time clinical feature categories under the supervision of three expert clinicians in this relevant field. Furthermore, the identified 25 features of Table 4.9 have been discussed with three healthcare specialists and according to them, the selected criteria have been rightly regarded to be the crucial predictive attributes in terms of practical PCOS identification (**empty citation**). This investigation shows that the PCA technique’s minimal yet optimal number of features can not only be used to deliver the best performance with ML classifiers, but can also be effectively utilised to implement a real-time autonomous PCOS detection model in the future.

Therefore the results of performance analysis indicates that, the machine learning model employing the proposed stacking ensemble method with five classifiers (SVM,LR,DT,KNN,NB) as base models and 'GradBoost' classifier as meta learner; utilizing the top 25 attributes from the dataset selected through PCA feature selection technique has been explored to be the highest performing classification model with 95.7% accuracy, 95.2% precision, 95.2% recall and 95.0% F1-score that outperforms all other models to classify PCOS and non-PCOS criteria.

4.3 Chapter Summary

In this article, three types of ensemble machine learning strategies (bagging, boosting and stacking) with multiple classifiers have been explored, trained and tested along with traditional machine learning techniques to classify PCOS and non-PCOS data. Most of the previous studies in this area were based on traditional ML classifiers. However, recently a few researchers have focused on applying ensemble techniques in PCOS detection, but their exploration techniques are based on typical bagging, boosting or voting type of ensemble models (Danaei Mehr and Polat 2022a; Bharati, Podder, M. Mondal, et al. 2022b). To the best of our knowledge, the proposed technique based on stacking ensemble classification approach where both traditional as well as boosting or bagging ensemble models are aggregated to provide a stronger prediction is a unique solution in this domain. Here in the stacked ensemble architecture, five types of weak traditional ML classifiers are used as the base models and then their predictions are integrated in a stronger meta-learner classification model to provide the final prediction. One from five types of boosting or bagging classifier has been used as the meta learner in the proposed stacked ensemble model to explore the best performing model where the highest performance has been acquired with 95.7% accuracy which is also higher than previous studies employing identical dataset. For instance, Bharti et al. (Bharati, Podder, M. Mondal, et al. 2022b) had acquired the best accuracy of 91.12% with voting ensemble technique, Nandipati et al. (Nandipati, Ying, and Wah 2020a) showed 93.12% accuracy with Random Forest classifier, Prapty et al. (Prapty and Shitu 2020a) acquired 93.5% accuracy employing Random Forest classifier and so on.

Furthermore, using feature selection strategies, the majority of previous studies randomly picked a specified number of features. For example, Bharti et al. (Bharati, Podder, and M. R. H. Mondal 2020) applied ML classifiers with ten statistically significant features based on p-values, Inan et al. (M. S. K. Inan, Ulfath, et al. 2021b) proposed to use most significant top twelve features, Danaei et al. (Danaei Mehr and Polat 2022a) had acquired best accuracy employing 28 features selected using Random Forest embedded feature selection technique and so on. However, hardly any study have investigated at how changing the numbers and combinations of features selected using that same feature selection method can affect the prediction result. Therefore, in this study, three distinct types of feature selection techniques (Chi-square, PCA and RFE) have been applied to identify the optimum features that are required for effective forecasting from the dataset's 40 attributes. Each of the feature selection techniques have been used to select different feature sets with top 35,30,25 and 20 attributes. And then the performances of the proposed stacking ensemble models as well as other traditional, bagging and boosting ensemble models are evaluated using those vast varieties of selected feature sets through performance metrics (accuracy, precision, recall and F1-scores). As per the findings of the comparative analysis, it has been observed that, the accuracy of most models using the feature set selected via Chi-square and RFE strategies improves up to the top 30 features and thereafter gradually diminishes, whereas in case of PCA feature selection approach the accuracy enhances upto top 25 features and then decreases. Therefore, comparing the performances of all the classifiers to categorize PCOS and non-PCOS patients, the result indicates that, the stacking ensemble model with 'Gradient Boosting' classifer as meta learner has outperformed other models utilizing the feature set of top 25 attributes picked using PCA technique. Furthermore, under the observation of expert clinicians, the highly prioritized 25 features selected using the PCA technique were sorted into real-time clinical categories.

CHAPTER 5

PCOS DETECTION USING USG IMAGE DATA

This chapter discusses the PCOS detection procedure with an extended machine learning technique using patients' USG image data. Firstly, the chapter includes the materials and methods for the proposed methodology, which has been subdivided into numbers of modules: image acquisition and pre-processing, image acquisition, applying machine learning classification techniques and performance analysis. Secondly, the chapter describes the exploration of machine learning techniques which includes the methodology of four types of ML techniques those are applied in this part of thesis: traditional ML, traditional ML with feature reduction, deep learning technique and finally the proposed extended ML technique. Finally, the comparative performance analysis with different types of performance metrics has been systematically discussed in the result analysis section of this chapter. However, the proposed methodology for this part of the thesis has been published in a reputed journal for PCOS prediction using USG image data (Suha and M. N. Islam 2022)

5.1 Phases of PCOS detection with USG image data

In this study, an extended machine learning classification technique has been proposed, trained and tested integrating ensemble ML models with Convolutional Neural Network(CNN) architecture that aims to differentiate between PCOS and non-PCOS ovaries. The ovary ultrasound images have been used as the input of the study from which the suggested method would determine whether or not the image reveals PCOS. Along with the proposed technique three other types of techniques have also been explored for classifying PCOS with an aim to conduct a comparative performance analysis of the suggested technique. The framework of the methodology used in this research is illustrated in Figure 5.1. The phases involved in this research are described briefly below.

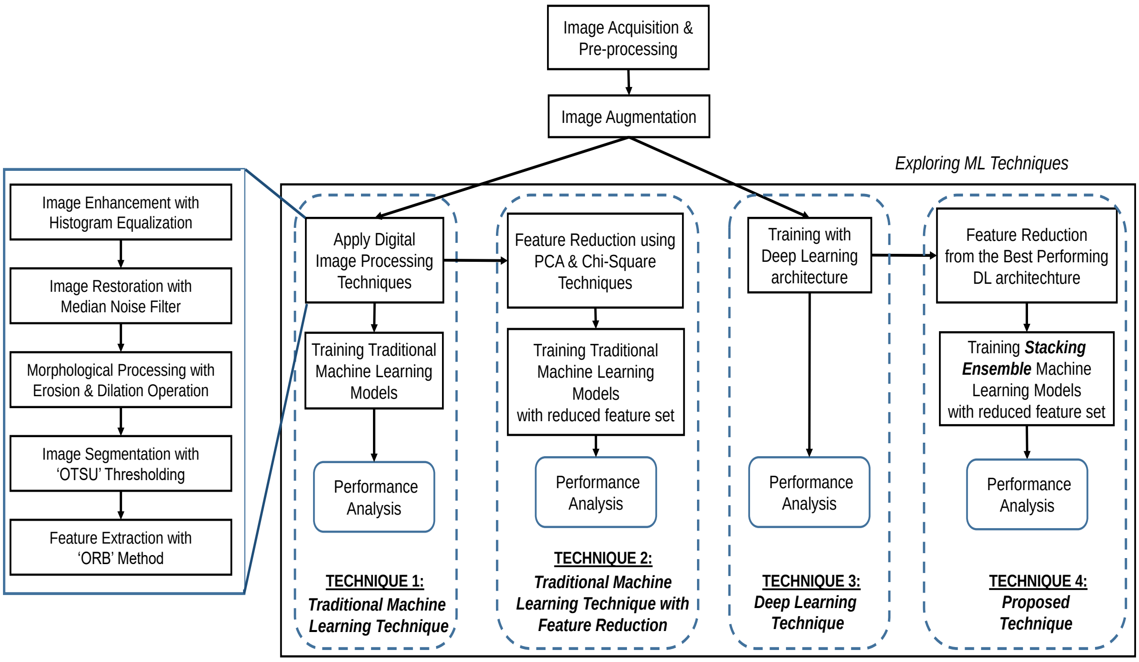


Fig. 5.1. Framework of research methodology

5.1.1 Image Acquisition and Pre-processing

The images that have been utilized in this study as inputs are ultrasonography images of the ovary. There are a total of 594 images among which 123 have been acquired from various open sources from the internet and the rest of them are collected from two diagnostic centers and three hospitals of Bangladesh including Combined Military Hospital (CMH), maintaining the ethical and privacy concerns. After consulting with two radiology specialists from Mymensignh Medical College Hospital, Bangladesh and Combined Military Hospital (CMH), Bangladesh; the images have been categorized to label as PCOS and Non-PCOS for using them as training dataset. Following the data labeling, it is observed that there are 306 photos with PCOS abnormality and 288 Non-PCOS images.

Because the images have been gathered from multiple sources, their formats and sizes differed, making them unsuitable to use in predictive models. Therefore, after acquiring the images from the repository to the implementation environment, they are converted to grayscale colorspace using an OpenCV python function 'COLOR_BGR2GRAY' and then all the images were resized to 224X224 size. Here, the Google Colaboratory has been used as a platform for implementation incorporating mainly the python scikit-learn and TensorFlow packages. At this stage, all of the images are stored in a 3D multichannel array with the shape (594, 224, 224, 3) that contains information about the images' plane, row, column, and channel. The labeling information or classes of each images are also stored in another 1D array which will be used as the target variable in predictive models.

5.1.2 Image Augmentation

Image augmentation is a powerful technique for reducing prediction error, which entails producing manipulated copies of images from the original training samples and therefore, depicting a more comprehensive set of potential data points (Shorten and Khoshgoftaar 2019). In this research, some basic image manipulations have been applied for image augmentation which includes geometric transformations like flipping, rotating and shifting; enhancing contrast; sharpening etc. Therefore, each of the input images have been augmented to three more images at this phase.

5.1.3 Applying Machine Learning Classification Techniques

Now, to train the machine learning models for classifying the images into PCOS and non-PCOS criteria; four types of technique have been conducted in this research (see Figure 5.1).The first technique involves the traditional approach of ML training where after applying relevant digital image processing phases the images are trained with conventional machine learning models. In the second technique the similar steps of the previous technique have been applied but with reduced set of features extracted from the images using two types (Chi-square and PCA) of feature reduction technique. The third technique has been to train the images with deep neural network (DNN) architecture where CNN technique have been applied incorporating different types of pre-trained models for transfer learning from which the best performing DNN architechture have been explored. Finally, the proposed technique have been applied where the reduced and significant features of the images have been acquired from the best performing deep learning architecture that has been explored from the previous technique and then using those reduced set of features a stacking ensemble machine learning model have been trained to classify the PCOS and non-PCOS images.

5.1.4 Performance Analysis

The predictive models' performances for each type of method are assessed on test data using several performance measures such as the Accuracy, Precision, Sensitivity (recall), Specificity, F1 score, execution time and AUC-ROC curve. Here, the execution time represents the computational time that is required for the algorithm to be trained. Other performance metrics are mainly based on the the comparative analysis of predicted values and true values of the training dataset which can be divided into four categories : true positive where the true value is positive and the predicted value is also positive; true Negative where the original value is negative, and also the predicted value has been negative; false positive in which the observation is negative, but the predicted value from the training has resulted to be positive and finally false negative that is the true value is positive, but is predicted to be negative. Basing on these evaluation the performance metrics can be written as following:

$$Accuracy = \frac{TruePositive + TrueNegative}{TruePositive + TrueNegative + FalsePositive + FalseNegative} \tag{5.1}$$

$$Precision = \frac{TruePositive}{TruePositive + FalsePositive} \tag{5.2}$$

$$Sensitivity(Recall) = \frac{TruePositive}{TruePositive + FalseNegative} \tag{5.3}$$

$$Specificity = \frac{TrueNegative}{TrueNegative + FalsePositive} \tag{5.4}$$

$$F - measure(F1score) = \frac{2 * (Precision * Sensitivity)}{Precision + Sensitivity} \tag{5.5}$$

Another evaluation metrics that has been employed in this study for assessing the performances of the models is the AUC (Area Under The Curve)- ROC (Receiver Operating Characteristics) curve. It is a familiar indicator of performance for classification methods at different threshold levels where AUC stands for the level or measurement of separability, and ROC is a probability curve. The curve is typically generated based on True Positive Rate (TPR). The higher the AUC, the more the model is effective at differentiating between patients with the condition and those who do not have. d False Positive Rate(FPR).

5.2 Exploring Machine Learning Techniques

The four different techniques that have been used in this study for classifying PCOS and Non-PCOS USG images are discussed hereafter.

5.2.1 Technique 1: Traditional Machine Learning Technique

In this technique, after performing the digital image processing operations to the images, they were fed into several conventional machine learning models to classify them. The procedure is briefly discussed below:

5.2.1.1 Performing Digital Image Processing Steps

For identifying the significant areas from the images to detect the follicles, the conventional digital image processing stages have been applied to the ovary USG images (see Figure 5.1: Image Processing Steps).

- The first phase is image enhancement, which is accomplished here using the histogram equalization method that improves the appearance and perception of information in pictures by providing a more consistently distributed grayscale histogram, resulting in a clearer image for viewers (Maini and Aggarwal 2010).

- The high impulsive noises in ultrasound images create severe uncertainty causing a detrimental impact on image interpretation for clinical diagnosis (Nadeem, Hussain, and Munir 2019). Thus, here a median filter technique is used as an image restoration step for noise reduction; which is a widely utilized nonlinear filter having outstanding edge preserving qualities with ability to reduce impulsive noise (Zhu and Huang 2012).

- After that, the required amount of morphological erosion and dilation has been done where dilation increases and erosion reduces the number of pixels on the edges of objects.

- The next phase is image segmentation to locate and highlight the follicles in the images for which the OTSU thresholding method has been emoloyed here. The OTSU technique is one of the most efficient automatic optimum global thresholding methods, which chooses a threshold value that maximizes the difference between pixel class variance and so divides the image into brighter and darker areas (Feng et al. 2017; Harb, Isa, and Salamah 2015).

- Finally, from the segmented images the key regions were extracted using Oriented FAST and Rotated BRIEF (ORB) feature extraction technique which is a faster and effective method for feature extraction to detect key points from the figure with enhanced computing efficiency as well as real-time benefits (Ma et al. 2020) and therefore the important areas are marked in the original images.

5.2.1.2 Splitting into Train and Test Data

After performing the above mentioned image processing steps the set of processed images each having size 224x224 are converted into a single dimension of using an OpenCV python function 'FLATTEN'. The two dimentional array now contains the pixel values of 594 images where each row indicates an input image with $224 * 224 = 50176$ columns or features containing the pixel values of that image. This array is then splitted into 70% train data and 30% test data to be utilized in the machine learning models.

5.2.1.3 Training Traditional Machine Learning Models

Ten different machine learning models have been applied to the train and test image dataset. The prediction type here is classification since the forecasting is a binary outcome to determine whether an USG image has PCOS or not. The seven models can be categorised as classical machine learning classification models (Menger, Scheepers, and Spruit 2018) that includes Naive Bayes model, Decision Tree model, Support Vector Machine(SVM), K Nearest Neighbour(KNN) model, Naive Bayes (NB) model; bagging ensemble machine learning models (H. Kaur, Malhi, and Pannu 2020) including Random Forest Classifier and boosting ensemble machine learning including gradient boosting, eXtreme Gradient Boosting (XGBoost) model, Adaptive Boosting(AdaBoost) classifier and Categorical Boosting(CATBoost).

5.2.1.4 Ablation Experiment:

In machine learning, ablation commonly implies the absence of a certain element or task from an AI system. Ablation studies shed light on the relative contributions of various structural and regularization elements to the efficiency of machine learning models. An ablation experiment examines the performance of an AI system by eliminating specific components in order to determine how those components affect the system as a whole. (Sheikholeslami et al. 2021; Sharif et al. 2018). In this case, in order to assess the relative significance of image processing approaches, an ablation study has been carried out by eliminating the image pre-processing stages and running the machine learning models on the raw images.

5.2.2 Technique 2: Traditional Machine Learning Technique with Feature Reduction

In the previous technique, it has been observed that, the input image array size was $594X50176$ which means each row of image consists of 50176 features containing pixel values. Such enormous amount of features may cause an over-fitting problem, lowering the models' prediction performance. In that circumstance, feature selection strategies can reduce the number of features from the dataset by selecting the most significant ones, thereby avoiding the curse of dimensionality (Kondo et al. 2019). Therefore, the second technique that has been performed in this study applied two types of feature selection approaches to find out the optimal and reduced set of features from the image dataset after conducting the image processing steps. And then those two types of dataset with reduced features are fed to the traditional machine learning models.

5.2.2.1 Feature Reduction using PCA and Chi-Square Technique

The chi-square feature selection strategy, which is one of the most common and useful feature selection techniques used in machine learning, is the first feature reduction method employed here. It is a numerical analysis that estimates departure from the expected distribution and prioritizes features by studying the connection between them (Thaseen and C. A. Kumar 2017). Another technique used for feature reduction in this study is the Principal component analysis (PCA) method, which is an efficient dimension reduction tool for feature prioritizing via quantitative simulation (Banerjee, Gupta, and J. Saha 2018). To achieve the feature reduction goal, PCA maps and reconstructs the original n-dimensional features to the needed k-dimensional features, where the k-dimensional features are new orthogonal features referred to as principal components (Zhao et al. 2019). Each of these strategies selected the most significant 25000 features from the 50176 features using their own methodology, resulting in two sets of datasets with reduced features at this step.

5.2.2.2 Training Traditional Machine Learning Models with reduced feature sets

The ten machine learning models specified in the first technique are trained again here with the dataset that have been divided into 70% train and 30% test data; where this time the models have utilized the two kinds of dataset with reduced set of features obtained by PCA and the Chi-Square feature selection techniques.

5.2.3 Technique 3: Deep learning Technique

After completing the image acquisition and augmentation phases, the third method performed in this study is the deep learning technique with Convolutional Neural Networks(CNN) (M. N. Islam, Aadeeb, et al. 2022). Figure 5.2 depicts the deep learning architecture used here.

In this study, a modified CNN model containing multiple layers with fine-tuning have been employed to classify the USG images. The techniques used in the multiple layers of the architecture have been described hereafter.

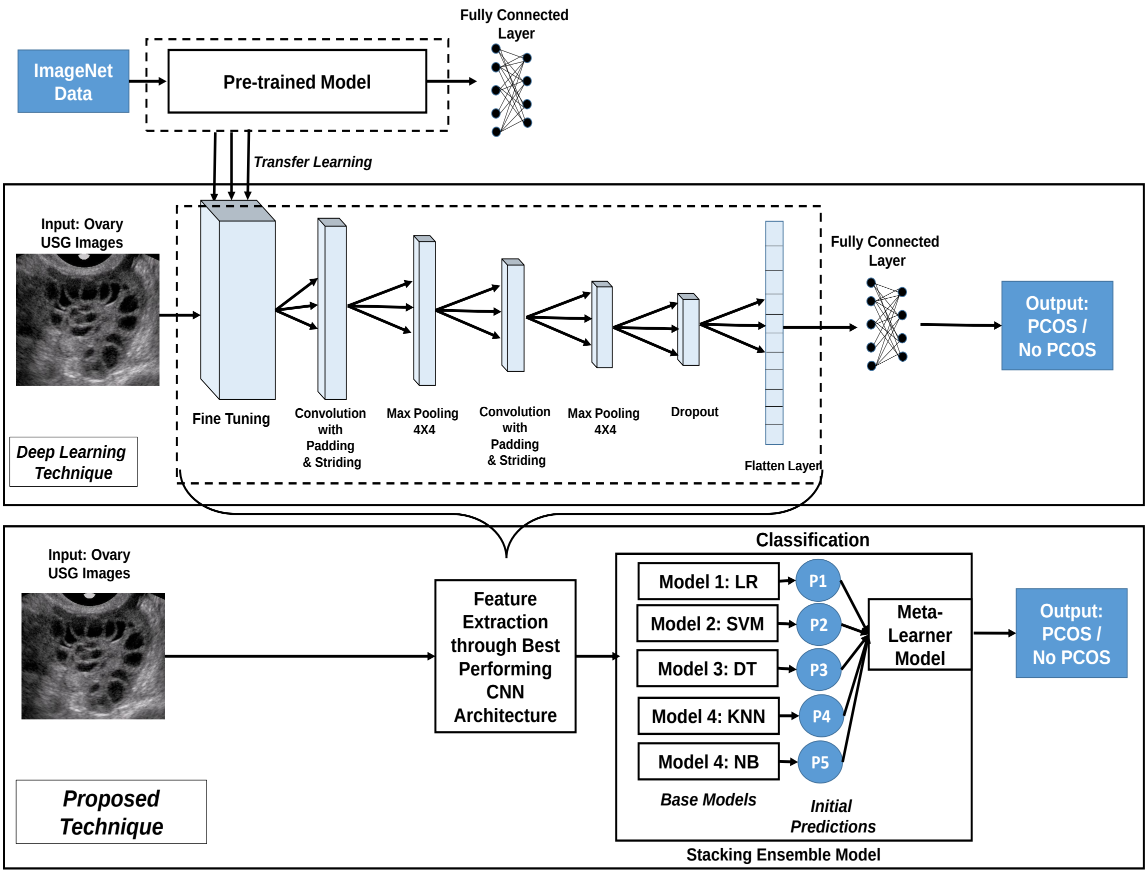


Fig. 5.2. Architecture of deep learning technique and proposed technique

- **Fine Tuning Layer:** The initial layer of the architecture is based on transfer learning with fine tuning that is a promising technique of machine learning which seeks to improve target learners' performance on target domains by transferring knowledge from different but relevant domains via pre-trained models (Zhuang et al. 2020). This study employs four types of best performing and popular pre-trained models, excluding their dense or fully connected layers for transfer learning; and then examines the combination of which pre-trained model in the CNN architecture can provide the best performance in case of this research. The pre-trained models that has been ulilized in this study are :
  - VGGNet16 model (A. Rehman et al. 2020; Nanda, Ghai, and Pande 2022) with 16 layers, 144M parameters and the best accuracy being 90.1% with Imagenet Data;
  - Xception model (Chollet 2017) with 36 layers, 22M parameters and the best

accuracy being 94.5% with Imagenet Data;

  – InceptionV3 model (Xia, C. Xu, and Nan 2017) with 48 layers, 24M parameters and the best accuracy being 93.7% with Imagenet Data;

  – MobileNet model (Howard et al. 2017) with 28 layers, 4.2M parameters and the best accuracy being 89.5% with Imagenet Data;

However, here an ablation experiment has been conducted in this technique to assess the relative importance of transfer learning with pre-trained model for which the CNN model has been trained and tested eliminating this fine tuning layer.

- **Convolution Layer:** The following layer is a ”Convolution layer” which is the CNN’s core foundational component having a set of filters or kernels with typically a smaller size than the training image, whose parameters must be trained over time and convolve with the actual image (Mostafa and Wu 2021). In this case, the kernel size has been considered to be $7X7$. The convolution layer here uses padding and striding techniques to provide a more accurate result, with padding providing one extra layer to the outer picture and striding handling the space between two successive kernel positions c. In this CNN architecture the value of padding is ’same’ which implies padding the input image with zeros uniformly to the outer side; and stride is 1 which means after the convolution using the kernel the output size will be same as the input size (Suha and Sanam 2022a). Now, in deep neural network, an activation function is being utilized to feed a weighted sum of input signals through, and the outcome from it is used as an input towards the following layer. In this case, ’Sigmoid’ activation function has been used which provides an output between 0 and 1 representing the probability classification outcome. The mathematical representation of the sigmoid function has been shown below:

$$Sigmoid\ Activation\ Function, f(x) = \frac{1}{1+e^{-(x)}} \tag{5.6}$$

  Similar kind of convolution operation has been conducted again in this CNN architecture after performing the following ’Pooling’ operation.

- **Pooing Layer:** The pooling layer is the next layer in the CNN architecture in this study which gradually reduces the spatial size of the image keeping the significant information in order to minimize the number of parameters and computations in the neural network (Yamashita et al. 2018). A ”Max Pooling” operation has been performed here, which yields the maximal value for each patch of the feature space with a pooling window size being $4x4$. Similar kind of max-pooling operation has been conducted again after the second convolution operation in the architecture.

- **Dropout Layer:** The next layer is the ”Dropour Layer” which helps to minimize overfitting problem as well as accelerate the training process as it changes input values to 0 randomly with a fixed frequency rate at every step throughout training phase (Nandini, A. S. Kumar, and Chidananda 2021). In this CNN model, the dropout rate has been considered to be 0.5 to avoid the overfitting problem.

- **Flatten Layer** The ”Flatten layer” is the subsequent layer, which turns the previous layer’s multi-dimensional output into a single dimension (M. N. Islam, Aadeeb, et al. 2022). A one-dimensional array is formed as a result of this layer’s output, which aids in the construction of the classifying neural network’s input layer, where the elements of the array is supplied towards each neuron. However, after performing all the previous CNN operations, the one dimensional array that has been generated in this phase possesses the most significant and reduced set of attributes from the input images. Thus, from the ”Fine tuning layer” to the ”Flatten layer” is considered to be the feature extraction segment of this deep learning technique.

- **Fully Connected Layer:** The ”Fully Connected Layer” or ”Dense Layer” is the last layer of the CNN design, which serves as the classifier layer in this technique. This layer, which is a sort of feed-forward artificial neural network, is placed at the bottom of the CNN model and within this layer, every neuron is linked to all neurons of the preceding layer, following the fundamental approach of the traditional multiple-layer perceptron neural network (Alzubaidi et al. 2021).

As the problem in this study is a binary classification task, the CNN model has been compiled utilizing the ’binary cross-entropy’ loss function and ‘adam’ optimizer with considering ‘accuracy’ as the core evaluation metrics for each classifiers per epoch. Therefore, employing the above mentioned CNN architecture the machine learning has been executed in 30 epochs in order to train the model using the input images and provide the categorization output. However, in deep learning method there is little necessity of processing the images before training, because the internal neural network processes the image to extract prominent features for classification.

5.2.4 Technique 4: Proposed Extended Technique

This proposed technique is the combination of both deep learning approach and stacking ensemble machine learning models; where the dominant features from the input dataset have been extracted using deep learning technique’s feature extraction segment and then the stacking ensemble machine learning models are incorporated to perform the classification phase utilizing that dataset containing reduced features. The block diagram of the proposed technique has been illustrated in Figure 5.2.

5.2.4.1 Feature Extraction from Best Performing DL Model

The internal neural networks of the deep learning architecture extract significant features from the images and the ”Flattten layer” generates a single feature vector, accumulating the output of the previous layers. Therefore, after training the deep learning model with the input image dataset, the outcome obtained from flatten layer with a one dimensional array will represent the reduced set of dataset containing the vital features. As one of the core benefits of CNN architecture is that it automatically detects significant attributes from the input image performing various layer of operations without the need for human intervention or image processing; thus, the proposed technique has extracted the reduced and optimal set of features from the best performing CNN model explored from the previously explained deep learning technique in Section 4.3.

5.2.4.2 Training Stacked Ensemble Machine Learning Models with Reduced Feature set

A stacking ensemble based ML classification approach has been applied here for classifying the PCOS or non-PCOS criteria in this proposed technique that offers several advantageous perspectives, such as: (a) it analyzes heterogeneous weak classifiers as well as learns them in parallel; (b) aggregates base classifiers by training a meta-learner to produce a stronger prediction based on the forecasts of the individual weak learners; and (c) thus, it minimizes variance and also enhances predictive force of the learning process. The stacking ensemble machine learning model has been utilized as the classifier phase in this proposed technique to replace the fully connected layer in the deep learning technique. From the best performing CNN model, the reduced set of features has been explored and then utilized as the input training data for the stacked machine learning classifiers.

Here, in the proposed stacked ensemble model, the dataset with reduced feature set is initially sent to the base learners. At this phase, the five types of widely utilized traditional machine learning classifiers have been considered to be the weak learners or base classifiers at level 0 of the stacked model, which are: Logistic Regression(LR), Support Vector Machine(SVM), Decision Tree(DT), K-Nearest Neighbour(KNN) and Gaussian Naive Bayes(NB) classifiers. Each of these base models gets trained independently using their respective prediction algorithms, resulting in forecasts symbolized p1, p2, p3, p4, and p5 in Figure 5.2. Following that, the predictions acquired from the level 0 models are fed into level 1 where a single classification model, or meta-learner, learns to generate final prediction from it. The meta-learner has been created at level 1 employing one stronger machine learning classifier. Here, while keeping the same base models at level 0, five types of classifiers have been explored as meta learners in level 1 which resulted in five varieties of the proposed architecture with an aim to explore the best performing one. The meta learner that has been used here is one of the five types of bagging or boosting classifiers (Random Forest, Adaptive Boosting, Gradient Boosting, CAT Boosting and XGBoosting) that is ultimately trained on top of level 0 to generate the final output based on the predictions returned by the base models. Thus, an extended stacked ensemble ML classifier has been proposed, trained and evaluated incorporating five types of traditional classifiers as base models and one boosting or bagging type of classifier as meta learner with an aim to differentiate between PCOS and non PCOS patient ovary USG images. The pseudocode of the proposed technique for PCOS detection has been shown in Algorithm 2.

Algorithm 2 Proposed Extended Technique for PCOS Detection

Input: $USGimg$: Ovarian USG Images;
$numF$: Number of reduced features
RF_{numF}: Reduced feature set with '$numF$' number of features
Output: list res: PCOS Detected (1), PCOS Not-Detected(0) [1] **forall** i in $USGimg$ **do**
end
Extract $numF$ features from Flatten_layer_CNN($numF$) of CNN model $RF_{numF} \Leftarrow Flatten\_layer\_CNN(numF)$ Use RF_{numF} feature set for PCOS prediction using $Stacked\_Ensemble\_ML\_classifier(RF_{numF})$ $res[i] \Leftarrow Stacked\_Ensemble\_ML\_classifier(RF_{numF})$ **return** res

5.3 Result Analysis

The following sections have been organized to present the findings from the interpretation of research results.

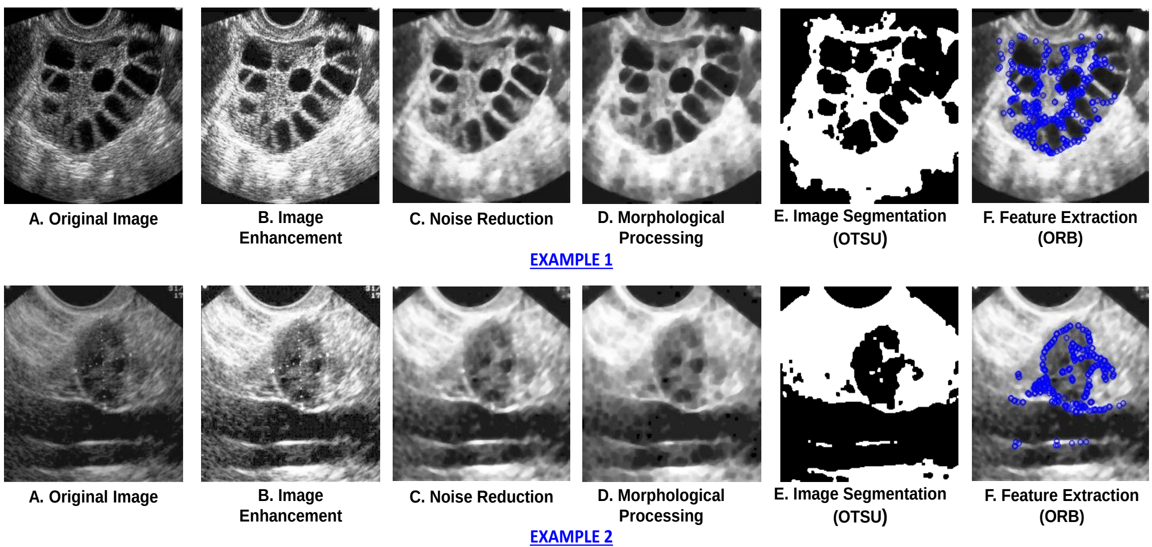


Fig. 5.3. USG Scans with Image Processing Steps

5.3.1 Findings from Machine Learning Techniques

Employing technique 1, the performances of the ten types of traditional machine learning models using several performance metrices and their execution times have been illustrated

Table 5.1: Performance Analysis of Test Data using Traditional Machine Learning Technique with and without image pre-processing (Technique 1)

ML Models	Without Image Pre-processing					With Image Pre-processing				
	Accu.	Prec	Rec.	F1-s	Time	Accu.	Prec.	Rec.	F1-s	Time
Log Reg	0.739	0.739	0.74	0.74	5.85	0.756	0.754	0.76	0.76	4.3
SVM	0.762	0.762	0.76	0.74	370.7	0.807	0.808	0.81	0.80	268.2
DT	0.714	0.715	0.72	0.71	24.8	0.737	0.773	0.73	0.75	10.4
KNN	0.718	0.719	0.72	0.72	13.9	0.849	0.870	0.85	0.85	9.1
NB	0.714	0.716	0.71	0.72	6.3	0.760	0.761	0.77	0.76	4.2
RF	0.778	0.779	0.77	0.77	8.2	**0.883**	0.883	0.88	0.88	5.8
GB	0.791	0.792	0.80	0.82	44.3	0.855	0.856	0.86	0.85	31.8
Ada Boost	0.718	0.716	0.71	0.71	272.2	0.815	0.813	0.82	0.81	218.9
XG Boost	**0.806**	0.810	0.80	0.81	172.2	0.856	0.855	0.85	0.86	150.1
CATBoost	0.712	0.710	0.72	0.70	72.2	0.790	0.792	0.79	0.79	74.1

in Table 5.1. In this technique, after image acquisition and augmentation phases, all the images are being processed using image processing techniques. The examples of USG scan with image processing technique is shown in Figure 5.3. The figure makes it apparent that image pre-processing steps change the original images substantially clearer while also allowing for the identification of the ovarian cysts that were barely apparent in the original image in the impacted locations. The models are then trained using the enhanced pictures with feature extraction containing 50176 features of pixel values, which have been collected at the completion of image processing steps. Here, for conducting ablation study to analyze the relative importance of image processing steps, the models have been trained and tested firstly without image pre-processing and then with image pre-processing. From the performance analysis in Table 5.1, it can be observed that, for all the classifiers, the performances significantly enhances after applying image-preprocessing steps. In case of training the models without image pre-processing, the best performance has been acquired in XG-Boosting model with 80.6% accuracy. On the other hand, employing image pre-processing techniques, Random Forest classification model outperforms all other classifiers in terms of performing with highest accuracy (88.3%) in less computation time (5.8 seconds) whereas the Naive Bayes classifier provides the least accuracy (76.0%) and SVM classifier takes the highest execution time (368.2 seconds). The AUC-ROC curve of the analysis without image pre-processing and with image pre-processing has been shown in Figure 5.6 (a) and

Table 5.2: Performance Analysis of Test Data using Traditional Machine Learning Technique with Feature Reduction (Technique 2)

ML	Using Chi-Square Feature Set					Using PCA Feature Set					
	Accu.	Prec	Rec.	F1-s	Time	Accu.	Prec.	Rec.	Spec	F1-s	Time
Log Reg	0.768	0.769	0.76	0.77	3.7	0.760	0.761	0.76	0.76	0.76	3.9
SVM	0.823	0.820	0.82	0.82	111.9	0.776	0.770	0.77	0.77	0.77	113.9
DT	0.639	0.647	0.64	0.63	14.6	0.637	0.673	0.63	0.54	0.65	10.2
KNN	0.778	0.822	0.73	0.72	6.3	0.782	0.819	0.77	0.72	0.77	7.6
NB	0.698	0.699	0.69	0.69	2.75	0.697	0.664	0.67	0.65	0.66	3.3
RF	**0.888**	0.886	0.89	0.88	4.3	0.793	0.781	0.72	0.70	0.72	4.8
GBoost	0.853	0.859	0.86	0.85	35.2	**0.805**	0.809	0.81	0.81	0.80	36.8
AdaBoost	0.773	0.789	0.77	0.77	177.2	0.790	0.792	0.79	0.78	0.79	104.1
XGBoost	0.783	0.779	0.77	0.77	75.2	0.765	0.769	0.76	0.74	0.76	86.8
CATBoost	0.811	0.813	0.81	0.81	48.3	0.799	0.795	0.79	0.79	0.80	50.5

Table 5.3: Performance Analysis of Test Data using Deep Learning Technique (Technique 3)

Deep Learning Models	Accuracy	Precision	Recall	Specificity	F1-Score	Time (s)
CNN Without Transfer Learning	0.7479	0.75	0.78	0.74	0.75	45.9
CNN with VGGNet16	**0.9780**	0.97	0.96	0.96	0.97	82.6
CNN with Xception	0.9187	0.92	0.91	0.92	0.91	96.7
CNN with InceptionV3	0.7625	0.76	0.75	0.76	0.76	55.6
CNN with MobileNet	0.9162	0.91	0.92	0.91	0.91	70.8

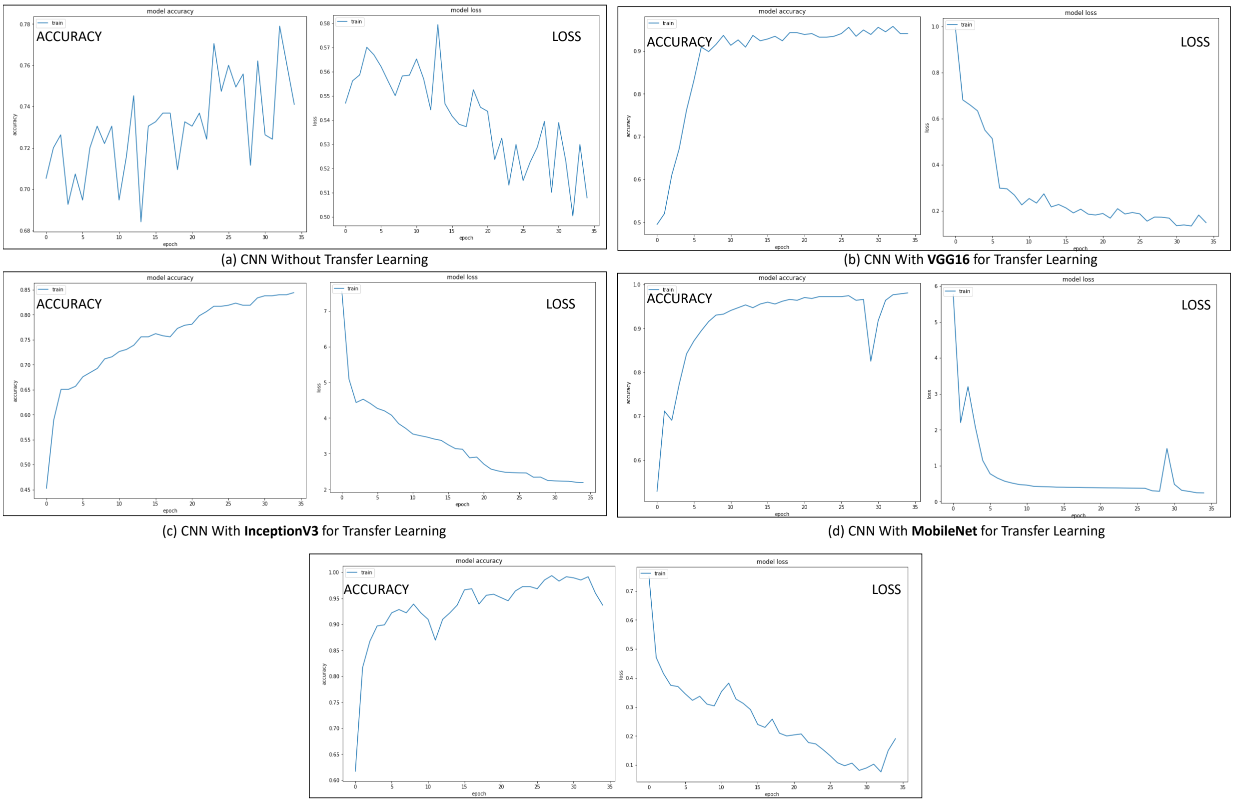
(b). Thus, the performance analysis clearly reveals that, image pre-processing techniques plays a significant role in acquiring good performances with machine learning classifiers. In the second type of technique, the two datasets with reduced features of 25,000 pixel columns acquired from the two types of feature reduction algorithms are applied to the traditional machine learning classifiers. The performances and execution times of the ten types of traditional machine learning models with both type of feature sets are shown in Table 5.2. From the table, it is apparent that after feature reduction the execution time for each of the models reduces significantly. It can be also observed that, most of the models perform better when employing chi-square feature sets than PCA technique. However, the Random Forest classifier outperforms the others, with an accuracy of 88.8% using the chi-square approach and Gradient Boosting model performs comparatively better using PCA feature set with 80.5% accuracy.

Table 5.4: Performance Analysis of Test Data using Proposed Technique (Technique 4)

	Base Models	**Meta-Learner**	Acc	Prec	Rec	Spec	F1-Sc	Time
Feature Extraction	LR	**Random Forest**	0.9844	0.98	0.98	0.97	0.98	0.09
with CNN	SVM	**XGBoost**	**0.9989**	0.99	1.00	1.00	0.99	0.05
& Classification	DT	**AdaBoost**	0.9959	0.99	0.98	0.99	0.99	1.07
with Stacking	KNN	**GradBoost**	0.9922	0.99	1.00	0.99	0.99	0.09
Ensemble Model	NB	**CATBoost**	0.9958	1.0	0.99	0.99	1.0	1.12

The 3rd technique has followed deep learning architecture in which a variety of pre-trained model has been employed in the initial layer using the transfer learning (TL) approach. For analyzing the relative importance of this transfer learning layer, an ablation study has also been conducted in this phase where the CNN model has been trained and tested without the transfer learning layer. Each model has been trained using 30 epochs. Table 5.3 shows the performances of the deep learning technique on a test dataset without transfer learning layer as well as utilizing four different types of pre-trained models each one executing over 30 epochs. Figure 5.4 shows the accuracy and loss per epoch for each type of the dnn techniques. Figure 5.6 (c) illustrates the AUC-ROC curve employing various forms of DNN technique. Here, from the performance analysis it is evident that all the deep learning approaches have gained significantly higher accuracy with the ”VGGNet16” pre-trained model surpassing the others, obtaining 97.80% accuracy in 82.7 seconds execution time. It is also noticeable that, the CNN model without transfer learning layer performs much worse with only 74.79% accuracy which is much lesser than the models employing transfer learning. This analysis reveals that, transfer learning plays a key role in efficient prediction of the classes. However, the performance analysis from 5.3 also shows that though the accuracy of the models using this dnn technique is much higher but the execution time for prediction is also very high which can be considered as a drawback of this technique.

Finally, in the proposed extended technique, the input dataset for the machine learning classifiers is the flatten layer’s output feature vector from the deep learning architecture incorporating the best performing CNN architecture with ”VGGNet16” pre-trained model used for tranfer learning. The extracted dataset with reduced features now contains only the 32 most significant attributes from the USG ovary images gained by the deep learning method’s



(a) CNN Without Transfer Learning

(b) CNN With **VGG16** for Transfer Learning

(c) CNN With **InceptionV3** for Transfer Learning

(d) CNN With **MobileNet** for Transfer Learning

(e) CNN With **Xception** for Transfer Learning

Fig. 5.4. Accuracy and Loss per epoch for DNN models (technique 3) (a) CNN without transfer learning (b)CNN with VGGNet16 for transfer learning (c) with InceptionV3 for transfer learning (d) with MobileNet for transfer learning (e) with Xception for transfer learning

neural networks. This dataset with extracted features has been fed to the proposed stacking machine learning classifiers where all the stacked ensemble model contains same base learners but five different types of meta-learners and therefore their comparative predictive performances are analyzed using performance metrics, as shown in Table 5.4. The results reveal that, the performances of all of the models have improved dramatically, with the XGBoost classification model as meta-learner achieving the greatest accuracy of 99.89% in just 0.05 seconds of execution time. Here, from this analysis it is apparent that, for all the varieties of the proposed technique the accuracy is always more than at least 98% with comparatively very less execution times. The AUC-ROC curve of the proposed technique has been shown in Figure 5.6(d), which also shows significantly efficient AUC scores than the other types of predictive models.

However, here the best performing meta-learner ”XGBoost” classifier is an efficient open-source boosting type of ensemble machine learning model. The model is trained here by

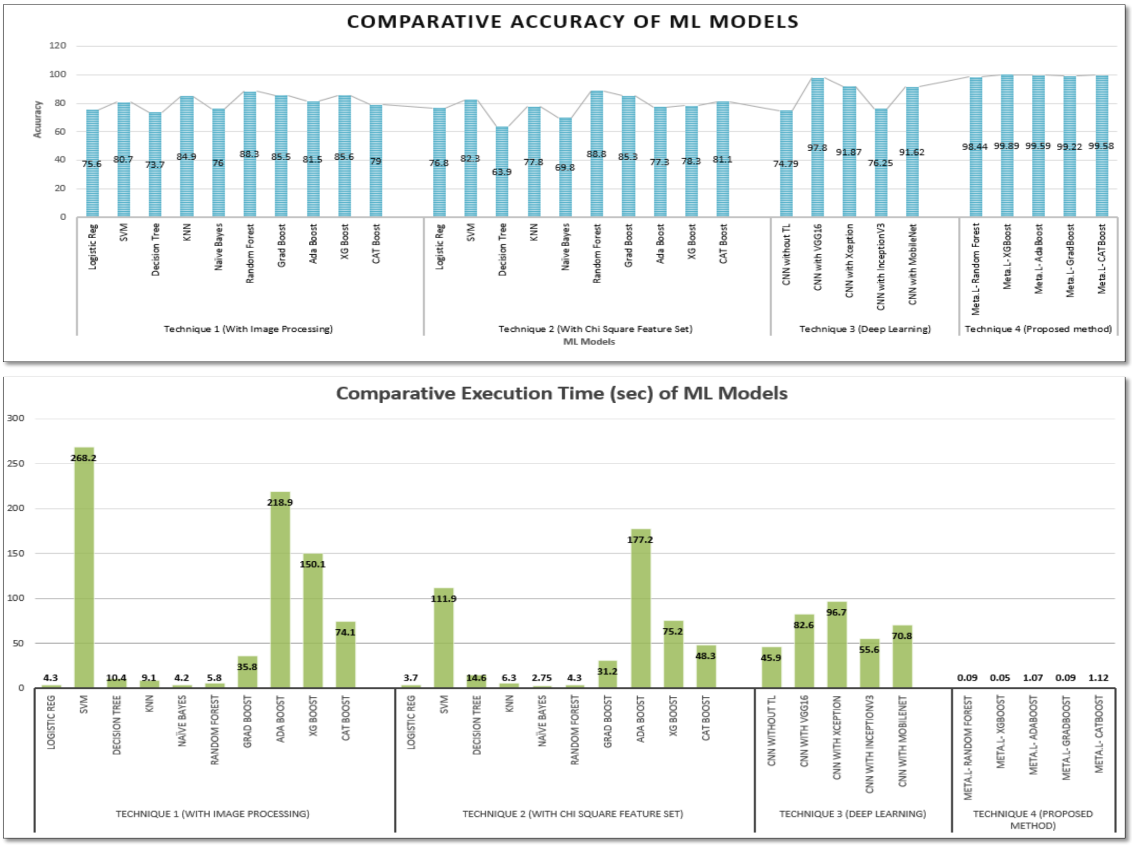
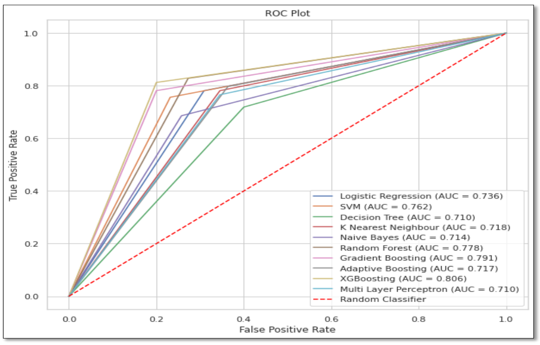
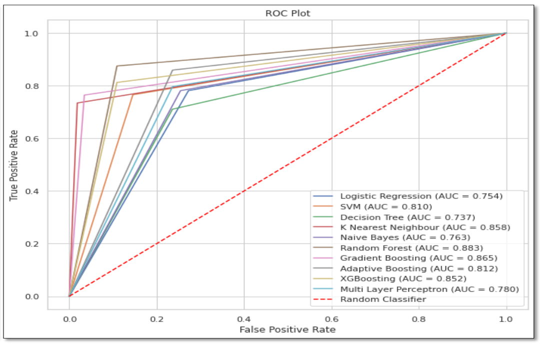


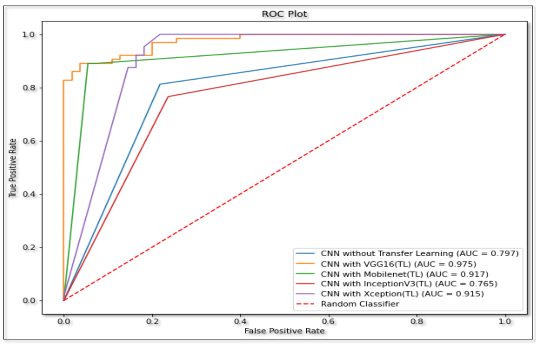
Fig. 5.5. Comparative analysis of (a) accuracy and (b) execution time for ML models using different techniques



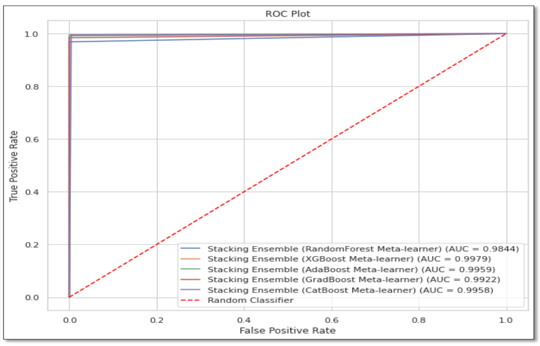
(a) AUC-ROC curve without image pre-processing



(b) AUC-ROC curve with image pre-processing



(c) AUC-ROC curve with deep learning technique



(d) AUC-ROC curve with proposed technique

Fig. 5.6. AUC-ROC curve for ML models (a)without image pre-processing (technique 1) (b) with image pre-processing (technique 1) (c) with deep learning (technique 3) (d) with proposed technique (technique 4)

minimizing the loss of an 'objective' function against the dataset with an aim to minimize the error, where the appropriate choice of loss function in XGBoost model is a critical hyperparameter to acquire desired performances. As the problem in this article is a binary classification task with two class labels, thus in the proposed technique the loss function used for predicting probabilities is 'objective=binary:logistic'.

5.3.2 Comparative Analysis

The comparative performance analysis of the accuracy and execution times of various machine learning models employing the four types of techniques in this study for PCOS detection are represented graphically in Figure 5.5. Also, the AUC-ROC curve of the techniques have been illustrated in Figure 5.6. The comparative analysis of the four methods for detecting PCOS from ultrasound images using machine learning techniques reveals that, the performances of the ten types of ML models employing traditional approach in technique 1 with all input features from the images are much poorer in terms of all the performance metrics (see Table 5.1). Though the performances of the models enhances after applying image pre-processing steps, but still the highest performance can be achieved in this technique with an accuracy of 88.3% in 5.8sec execution time employing Random Forest classifier. The performances of the models enhance a little bit when numerical feature selection techniques like chi-aquare and PCA have been applied to the dataset in technique 2, where the highest accuracy has been gained to be 88.8% with 4.3 sec execution time with Random Forest classifier trained with reduced feature set of chi-square technique (see Table 5.2).
However, the performances of detecting the PCOS get much higher when deep learning model is utilized for classification in technique 3 employing transfer learning technique, with an highest accuracy of 97.80% with VGGNet16 used as the layer of transfer learning in the proposed CNN architecture (see Table 5.3). But, the limitations of using deep learning models are that they are expensive to train and demand a lot of processing power and time. Therefore, as compared to typical machine learning models the execution time is significantly longer in deep learning models, which is evident in Table 5.3 (best performing model takes 82.6 sec execution time). To overcome these drawbacks, this study have

implemented a hybrid model in method 4, in which the feature extraction phase from the images has be conducted by the best performing deep learning model and then the stacking ensemble machine learning models are utilized to do the classification. This method yields substantially higher accuracy in a short amount of time (see Table 5.4). It is noticeable that, for each of the ML model the accuracy increases and execution time decreases dramatically with the proposed technique. For example, for Random Forest classifier the best acquired accuracy with technique 1 is 88.3% and with technique 2 is 88.8%; which enhances significantly providing 98.44% accuracy while employing as a meta-learner of the stacked ensemble model of the proposed technique. Also, as the classifiers predicts the classes based on only 32 dominant features of the USG images, the execution times are comparatively much lesser in the proposed technique. Therefore, the comparative performance analysis reveals that, the proposed technique employing CNN with ”VGGNet16” pre-trained model used for transfer learning as the feature extractor and the stacking ensemble model with ”XGBoost” classification model used for meta-learner as the classifier of this problem can categorize PCOS and non-PCOS USG images in not only with the best performances but also solve the problem in least execution time.

5.4 Chapter Summary

The presented work in this research can be a pioneer study in the domain of PCOS detection which has incorporated the advantages of both conventional and deep learning techniques. The findings of this study can be significantly beneficial towards both patients and healthcare providers in identifying PCOS quickly and efficiently. The suggested hybrid method has combined transfer learning technique using powerful pre-trained models in a CNN architecture to extract the most significant feature set from the images and then stacking ensemble machine learning technique have been employed to classify the images. To analyze the efficacy of the model and evaluate its performances, three types of other existing methodologies have also been performed in this study. The first two existing techniques for PCOS detection from USG images using traditional ML algorithms have appeared to be much inefficient as they provides significantly less accuracy as well necessitate tedious digital image processing steps for feature extraction and also requires another feature selection method for the reduction of massive amount of image features. Also, the third existing technique with only deep learning technique requires much higher computational space and time. Ultimately, after several analysis the experimental findings reveal that, the proposed hybrid strategy of employing the ”VGG16” pre-trained model for transfer learning in the CNN architecture for feature extraction and then the ”XGBoost” machine learning model as the meta-learner of stacking ensemble model for image classification yields the maximum accuracy of 99.89% with a relatively shortest execution time to detect PCOS from ultrasound images.

An effective machine learning approach to detect PCOS from ovarian ultrasound images is considered to be a potentially valuable solution for recovering thousands of women’s reproductive health. Most of the studies that had been conducted in this field (for example the works conducted in (Mandal, D. Saha, and Sarkar 2021; YILMAZ and ÖZMEN 2020; Gopalakrishnan and Iyapparaja 2020; Setiawati, Tjokorda, et al. 2015; Sitheswaran and Malarkhodi 2014; Mehrotra, Chakraborty, et al. 2011; Deng, Y. Wang, and P. Chen 2008; Rachana et al. 2021a; Nilofer et al. 2021a; Gopalakrishnan and Iyapparaja 2019; Purnama et al. 2015a; Deshpande and Wakankar 2014)) for PCOS detection are the implementation of various forms of digital image processing techniques for follicle detection and then applying conventional machine learning models for classification. Recently, a few researchers have also focused on applying deep neural network with CNN architecture to detect PCOS from medical images (for example the works conducted in (Vikas, Radhika, and Vineesha 2021; Cahyono, Mubarok, Wisesty, et al. 2017)). Yet, little or no research has been conducted in this area for PCOS diagnosis that uses both traditional ML classifiers and deep neural networks together with transfer learning and fine tuning. Moreover, even if some other research from different domains have adopted this type of integrated model (for example (Pang et al. 2019)), but hardly any study has been found that combined the CNN techniques with stacked ensemble machine learning classifiers with an aim to detect anomaly from medical images. Thus, to the best of our knowledge, the proposed technique in this study, where CNN with transfer learning and fine tuning utilized as the feature extractor and then stacked ensemble machine learning with the integration of classical and boosting ensemble models as the classifier is a unique and novel solution in the domain of medical image analysis for disease identification.

CHAPTER 6

CONCLUSIONS

This chapter provides the concluding remarks which is discussed into a number of modules. These are thesis outcome, thesis implications, thesis limitations and future work. The thesis outcomes are outlined in a structured manner firstly. Then, implications of this thesis are discussed. Following that, some limitations of this thesis are stated followed by a number of potential opportunities for future work.

6.1 Thesis Outcomes

A number of findings have been found from this research work which can be very helpful in the domain of automated PCOS detection. Firstly, the systematic literature review summarizes the existing relevant studies concentrating on the use of various computer-assisted methodologies for PCOS detection by thoroughly examining them with an aim to explore the state-of-art viewpoints, identify the shortcomings and determine the possible future study areas. Secondly, an extended machine learning classifier has been developed to predict the existence of PCOS in female body based on their symptom data. Here, among a vast numbers of attributes or symptoms, the minimal and most significant features are also explored in this study; as such the first research objective has been fulfilled. Thirdly, an extended hybrid machine learning technique for PCOS detection based on patients' USG image data has been developed in this research which combines the advantages of both the deep learning and the ensemble machine learning techniques; as such the second research objective has been achieved. Finally, to acquire the final research objective, the comparative performance analysis of the both kind of proposed techniques with the previously adopted ML techniques highlights the efficacy of the suggested methodologies in this domain for PCOS detection.

6.2 Thesis Contributions and Implications

In this study, the proposed technique with patient symptom data is based on stacking ensemble classification, where both traditional as well as boosting or bagging ensemble models are aggregated to provide a stronger prediction, which is a unique solution in this domain. The study also explores the optimum features in a rigorous manner to provide the best predictive output using the minimal attributes with the proposed technique. Moreover, the proposed hybrid technique for PCOS detection using USG images has combined transfer learning technique using powerful pre-trained models in a CNN architecture to extract the most significant feature set from the images and then stacking ensemble machine learning technique have been employed to classify the images. This technique combines the benefits of both neural network with transfer learning and fine tuning as well as ensemble machine learning for the predictive analysis which is also a novel solution in this domain. The comparative analysis of the proposed techniques with the existing ML techniques again proves its efficacy where using both types of data modalities the techniques outperforms.

Therefore, the methodology presented in this study can be a pioneer in effectively detecting PCOS from patients symptoms and test results through machine learning strategies and thereby can play a potentially beneficial role in improving the reproductive health of thousands of women. It is obvious that there are a vast array of potential areas for research in this domain, and the findings of this study will undoubtedly aid scholars in understanding the depth and breadth of existing studies, requirements and limitations to conduct more experiments and studies for PCOS detection in future. The findings of this study can be significantly beneficial towards both patients and healthcare providers in identifying PCOS quickly and efficiently combining the advantages of multiple machine learning classifiers ensembled in one robust model employing minimal number of attributes and thus it is anticipated to be widely used in the real-world clinical practices. Moreover, the findings will greatly benefit healthcare practitioners in identifying PCOS using ultrasound images more quickly and accurately combining the advantages of both traditional machine learning and deep learning approaches and thus it is anticipated to be widely used in the real-world scenario.

The study's outcome can be effectively helpful for the physicians in the arduous task evaluating patients by simplifying the complex diagnostic procedure of PCOS. This computational technique can be deployed in the healthcare facilities of rural areas to detect PCOS autonomously where there is scarcity of expert physicians and resources. Additionally, the literature review that has been conducted here also will aid the scholars and healthcare professionals to address the usefulness of different computer-assisted techniques, such as machine learning, image processing etc. in real-time decision-making for illness diagnosis at an early stage.

6.3 Thesis Limitations

The research has few limitations as well which are stated below:

- As the study depends on the clinical data of the patients, the core challenge of this study was to collect dataset. Collecting medical data or images of the patients in the least developing country is quite challenging and as a result the machine learning techniques have been applied on a limited number of images (594 images, 541 patient symptom data) due to the lack of a dataset.
- The dataset which has been used to implement the proposed methodology is an open access data, the researchers of this thesis could not directly real-time data from any hospital due to numbers of obstacles.
- Owing to a lack of vast dataset, one of the study's flaws was that it only used machine learning algorithms on a small number of patient symptom data.
- As the work is related to healthcare analytics, it is important to conduct a systematic evaluation of the proposed technique in context of the doctor's perspectives. But, in case of this research work, such kind of evaluation has not been conducted.

- The proposed technique is based on deep learning feature extraction and stacked ensemble machine learning classifier which is quite difficult to make explainable and thus may seem like black box towards the physicians.
- The proposed technique can only be in use of patients and doctors, if it had been implemented as an interactive platform. But, no interactive software version has been designed or developed based on the proposed technique for practical use.

6.4 Future Work

There are a number of ways to enhance this research work in future. These are discussed as follows:

- This study can be incorporated with real-time dataset from a hospital from Bangladesh and then apply the proposed techniques for PCOS detection.
- The research work can be elaborated through the development of an interactive mobile application or website, so that the patients or doctors can use the prediction of PCOS practically.
- The results of the proposed predictive models can be evaluated more critically in context of the doctors perspective to analyze its efficacy clinically.
- For intelligent clinical applications, explainable AI plays an important role in providing an explanation alongside sufficient justification of AI system predictions. Therefore, PCOS detection process can be expanded more incorporating the techniques of eXplainable AI (XAI) with the current study.
- The proposed methodology can also be applied in other fields of clinical illness predictions for healthcare analytics.

REFERENCES

Abayomi-Alli, Olusola O et al. (2022). "An Ensemble Learning Model for COVID-19 Detection from Blood Test Samples." In: *Sensors* 22.6, p. 2224.

Abraham Gnanadass, Subeka, Yogamaya Divakar Prabhu, and Abilash Valsala Gopalakrishnan (2021). "Association of metabolic and inflammatory markers with polycystic ovarian syndrome (PCOS): an update." In: *Archives of Gynecology and Obstetrics* 303.3, pp. 631–643.

Acharya, U Rajendra, Filippo Molinari, et al. (2015). "Ovarian tissue characterization in ultrasound: a review." In: *Technology in cancer research & treatment* 14.3, pp. 251–261.

Acharya, U Rajendra, S Vinitha Sree, et al. (2013). "Ovarian tumor characterization and classification using ultrasound—a new online paradigm." In: *Journal of digital imaging* 26.3, pp. 544–553.

Adla, Yasmine A Abu et al. (2021). "Automated Detection of Polycystic Ovary Syndrome Using Machine Learning Techniques." In: *2021 Sixth International Conference on Advances in Biomedical Engineering (ICABME)*. IEEE, pp. 208–212.

Ajmal, Nida, Sanam Zeib Khan, and Rozeena Shaikh (2019). "Polycystic ovary syndrome (PCOS) and genetic predisposition: A review article." In: *European journal of obstetrics & gynecology and reproductive biology: X* 3, p. 100060.

Akbar, Wasif et al. (2020). "Development of Hepatitis Disease Detection System by Exploiting Sparsity in Linear Support Vector Machine to Improve Strength of AdaBoost Ensemble Model." In: *Mobile Information Systems* 2020.

Alqudah, Ali Mohammad (2019). “Ovarian cancer classification using serum proteomic profiling and wavelet features a comparison of machine learning and features selection algorithms.” In: *Journal of Clinical Engineering* 44.4, pp. 165–173.

Alzubaidi, Laith et al. (2021). “Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions.” In: *Journal of big Data* 8.1, pp. 1–74.

Anagnostis, Panagiotis, Basil C Tarlatzis, and Robert P Kauffman (2018). “Polycystic ovarian syndrome (PCOS): Long-term metabolic consequences.” In: *Metabolism* 86, pp. 33–43.

Arentz, Susan et al. (2021). “Perceptions and experiences of lifestyle interventions in women with polycystic ovary syndrome (PCOS), as a management strategy for symptoms of PCOS.” In: *BMC women’s health* 21.1, pp. 1–8.

Ashish, Ladda, Sravan Kumar, and Sahithi Yeligeti (2021). “Ischemic heart disease detection using support vector machine and extreme gradient boosting method.” In: *Materials Today: Proceedings*.

Azar, Ahmad Taher et al. (2014). “A random forest classifier for lymph diseases.” In: *Computer methods and programs in biomedicine* 113.2, pp. 465–473.

B, Vikas, Radhika Yalavarthi, and Vineesha K (Jan. 2021). “Detection of Polycystic Ovarian Syndrome using Convolutional Neural Networks.” In: *International Journal of Current Research and Review* 13, pp. 155–159. DOI: 10.31782/IJCRR.2021.13630.

Bahad, Pritika and Preeti Saxena (2020). “Study of adaboost and gradient boosting algorithms for predictive analytics.” In: *International Conference on Intelligent Computing and Smart Communication 2019*. Springer, pp. 235–244.

Balen, Adam H et al. (2003). “Ultrasound assessment of the polycystic ovary: international consensus definitions.” In: *Human reproduction update* 9.6, pp. 505–514.

Banerjee, Soumyendu, Rajarshi Gupta, and Jayanta Saha (2018). “Compression of multilead electrocardiogram using principal component analysis and machine learning approach.” In: *2018 IEEE Applied Signal Processing Conference (ASPCON)*. IEEE, pp. 24–28.

Bednarska, Sylwia and Agnieszka Siejka (2017). "The pathogenesis and treatment of polycystic ovary syndrome: What's new?" In: *Advances in Clinical and Experimental Medicine* 26.2.

Bharati, Subrato, Prajoy Podder, M Mondal, et al. (2022a). "Ensemble Learning for Data-Driven Diagnosis of Polycystic Ovary Syndrome." In: *International Conference on Intelligent Systems Design and Applications*. Springer, pp. 1250–1259.

— (2022b). "Ensemble Learning for Data-Driven Diagnosis of Polycystic Ovary Syndrome." In: *International Conference on Intelligent Systems Design and Applications*. Springer, pp. 1250–1259.

Bharati, Subrato, Prajoy Podder, and M Rubaiyat Hossain Mondal (2020). "Diagnosis of polycystic ovary syndrome using machine learning algorithms." In: *2020 IEEE Region 10 Symposium (TENSYMP)*. IEEE, pp. 1486–1489.

Boomidevi, R and S Usha (2021a). "Performance Analysıs of Polycystıc Ovary Syndrome (PCOS) Detectıon System Usıng Neural Network Approach." In: *Data Engineering and Communication Technology*. Springer, pp. 449–459.

— (2021b). "Performance Analysıs of Polycystıc Ovary Syndrome (PCOS) Detectıon System Usıng Neural Network Approach." In: *Data Engineering and Communication Technology*. Springer, pp. 449–459.

Brattain, Laura J et al. (2018). "Machine learning for medical ultrasound: status, methods, and future opportunities." In: *Abdominal radiology* 43.4, pp. 786–799.

Breiman, Leo (2001). "Random forests." In: *Machine learning* 45.1, pp. 5–32.

Cahyono, B, MS Mubarok, UN Wisesty, et al. (2017). "An implementation of convolutional neural network on PCO classification based on ultrasound image." In: *2017 5th International Conference on Information and Communication Technology (ICoIC7)*. IEEE, pp. 1–4.

Callahan, Alison and Nigam H Shah (2017). "Machine learning in healthcare." In: *Key Advances in Clinical Informatics*. Elsevier, pp. 279–291.

Chandel, Khushboo et al. (2016). “A comparative study on thyroid disease detection using K-nearest neighbor and Naive Bayes classification techniques.” In: *CSI transactions on ICT* 4.2, pp. 313–319.

Chandrasekar, Thaventhiran et al. (2022). “Lung cancer disease detection using service-oriented architectures and multivariate boosting classifier.” In: *Applied Soft Computing* 122, p. 108820.

Chen, Tianqi and Carlos Guestrin (2016). “Xgboost: A scalable tree boosting system.” In: *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pp. 785–794.

Chen, Wei et al. (2021). “Evaluation of different boosting ensemble machine learning models and novel deep learning and boosting framework for head-cut gully erosion susceptibility.” In: *Journal of Environmental Management* 284, p. 112015.

Chollet, François (2017). “Xception: Deep learning with depthwise separable convolutions.” In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1251–1258.

Condorelli, Rosita A et al. (2017). *PCOS and diabetes mellitus: from insulin resistance to altered beta pancreatic function, a link in evolution*.

Couto Alves, A et al. (2017). “Metabolic profiling of polycystic ovary syndrome reveals interactions with abdominal obesity.” In: *International Journal of Obesity* 41.9, pp. 1331–1340.

Danaei Mehr, Homay and Huseyin Polat (2022a). “Diagnosis of polycystic ovary syndrome through different machine learning and feature selection techniques.” In: *Health and Technology* 12.1, pp. 137–150.

— (2022b). “Diagnosis of polycystic ovary syndrome through different machine learning and feature selection techniques.” In: *Health and Technology* 12.1, pp. 137–150.

Dapas, Matthew and Andrea Dunaif (Jan. 2022). “Deconstructing a Syndrome: Genomic Insights Into PCOS Causal Mechanisms and Classification.” In: *Endocrine Reviews*. ISSN: 0163-769X. DOI: `10.1210/endrev/bnac001`. URL: `https://doi.org/10.1210/endrev/bnac001`.

Deif, MA, RE Hammam, and A Solyman (2021). “Gradient Boosting Machine Based on PSO for prediction of Leukemia after a Breast Cancer Diagnosis.” In: *Int. J. Adv. Sci. Eng. Inf. Technol* 11, pp. 508–515.

Deng, Yinhui, Yuanyuan Wang, and Ping Chen (2008). “Automated detection of polycystic ovary syndrome from ultrasound images.” In: *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, pp. 4772–4775.

Denny, Amsy et al. (2019a). “i-hope: Detection and prediction system for polycystic ovary syndrome (pcos) using machine learning techniques.” In: *TENCON 2019-2019 IEEE Region 10 Conference (TENCON)*. IEEE, pp. 673–678.

— (2019b). “i-hope: Detection and prediction system for polycystic ovary syndrome (pcos) using machine learning techniques.” In: *TENCON 2019-2019 IEEE Region 10 Conference (TENCON)*. IEEE, pp. 673–678.

Deshpande, Sharvari S and Asmita Wakankar (2014). “Automated detection of polycystic ovarian syndrome using follicle recognition.” In: *2014 IEEE International Conference on Advanced Communications, Control and Computing Technologies*. IEEE, pp. 1341–1346.

Devika, R, Sai Vaishnavi Avilala, and V Subramaniyaswamy (2019). “Comparative study of classifier for chronic kidney disease prediction using naive Bayes, KNN and random forest.” In: *2019 3rd International conference on computing methodologies and communication (ICCMC)*. IEEE, pp. 679–684.

Dewi, RM, UN Wisesty, et al. (2018). “Classification of polycystic ovary based on ultrasound images using competitive neural network.” In: *Journal of Physics: Conference Series*. Vol. 971. 1. IOP Publishing, p. 012005.

Escobar-Morreale, Héctor F (2018). “Polycystic ovary syndrome: definition, aetiology, diagnosis and treatment.” In: *Nature Reviews Endocrinology* 14.5, pp. 270–284.

Feng, Yuncong et al. (2017). “A multi-scale 3D Otsu thresholding algorithm for medical image segmentation.” In: *Digital Signal Processing* 60, pp. 186–199.

Freund, Yoav, Robert E Schapire, et al. (1996). “Experiments with a new boosting algorithm.” In: *icml*. Vol. 96. Citeseer, pp. 148–156.

Garg, Deepika and Reshef Tal (2016). "The role of AMH in the pathophysiology of polycystic ovarian syndrome." In: *Reproductive biomedicine online* 33.1, pp. 15–28.

George, Shemi and Athulya Alex (2021). "Assessment of Symptoms and Diet Intake in Young Adult with Polycystic Ovary Syndrome (PCOS)." In: *Journal of Scientific Research* 65.4.

Ghiasi, Mohammad M, Sohrab Zendehboudi, and Ali Asghar Mohsenipour (2020). "Decision tree-based diagnosis of coronary artery disease: CART model." In: *Computer methods and programs in biomedicine* 192, p. 105400.

Gopalakrishnan, C and M Iyapparaja (2019). "Detection of polycystic ovary syndrome from ultrasound images using SIFT descriptors." In: *Bonfring International Journal of Software Engineering and Soft Computing, 9 (2), 26* 30.

— (2020). "Active contour with modified Otsu method for automatic detection of polycystic ovary syndrome from ultrasound image of ovary." In: *Multimedia Tools and Applications* 79.23, pp. 17169–17192.

— (2021). "Multilevel thresholding based follicle detection and classification of polycystic ovary syndrome from the ultrasound images using machine learning." In: *International Journal of System Assurance Engineering and Management*, pp. 1–8.

Granitto, Pablo M et al. (2006). "Recursive feature elimination with random forest for PTR-MS analysis of agroindustrial products." In: *Chemometrics and intelligent laboratory systems* 83.2, pp. 83–90.

Han, Jiawei, Jian Pei, and Micheline Kamber (2011). *Data mining: concepts and techniques*. Elsevier.

Harb, Suheir M ElBayoumi, Nor Ashidi Mat Isa, and Samy A Salamah (2015). "Improved image magnification algorithm based on Otsu thresholding." In: *Computers & Electrical Engineering* 46, pp. 338–355.

Hartati, Sri, Aina Musdholifah, et al. (2019). "Machine learning assisted medical diagnosis for segmentation of follicle in ovary ultrasound." In: *International Conference on Soft Computing in Data Science*. Springer, pp. 71–80.

Hassan, Malik Mubasher and Tabasum Mirza (2020). "Comparative analysis of machine learning algorithms in diagnosis of polycystic ovarian syndrome." In: *Int. J. Comput. Appl* 975, p. 8887.

Heimerl, Florian et al. (2014). "Word cloud explorer: Text analytics based on word clouds." In: *2014 47th Hawaii international conference on system sciences*. IEEE, pp. 1833–1842.

Hiremath, PS and Jyothi R Tegnoor (2010). "Follicle detection in ultrasound images of ovaries using active contours method." In: *2010 International Conference on Signal and Image Processing*. IEEE, pp. 286–291.

Howard, Andrew G et al. (2017). "Mobilenets: Efficient convolutional neural networks for mobile vision applications." In: *arXiv preprint arXiv:1704.04861*.

Iatrakis, G et al. (2006). "Polycystic ovarian syndrome, insulin resistance and thickness of the endometrium." In: *European Journal of Obstetrics & Gynecology and Reproductive Biology* 127.2, pp. 218–221.

Inan, Muhammad Sakib Khan, Rizwan Hasan, and Fahim Irfan Alam (2021). "A hybrid probabilistic ensemble based extreme gradient boosting approach for breast cancer diagnosis." In: *2021 IEEE 11th Annual Computing and Communication Workshop and Conference (CCWC)*. IEEE, pp. 1029–1035.

Inan, Muhammad Sakib Khan, Rubaiath E Ulfath, et al. (2021a). "Improved sampling and feature selection to support extreme gradient boosting for PCOS diagnosis." In: *2021 IEEE 11th Annual Computing and Communication Workshop and Conference (CCWC)*. IEEE, pp. 1046–1050.

— (2021b). "Improved sampling and feature selection to support extreme gradient boosting for pcos diagnosis." In: *2021 IEEE 11th Annual Computing and Communication Workshop and Conference (CCWC)*. IEEE, pp. 1046–1050.

Isah, Omeiza Rabiu, AD Usman, and AM Tekanyi (2015). "A review on computer assisted follicle detection techniques and polycystic ovarian syndrome (PCOS) diagnostic systems." In.

Islam, Iyolita and Muhammad Nazrul Islam (2022). “Digital intervention to reduce counterfeit and falsified medicines: A systematic review and future research agenda.” In: *Journal of King Saud University-Computer and Information Sciences*.

Islam, Mohammed J et al. (2007). “Investigating the performance of naive-bayes classifiers and k-nearest neighbor classifiers.” In: *2007 International Conference on Convergence Information Technology (ICCIT 2007)*. IEEE, pp. 1541–1546.

Islam, Muhammad Nazrul, Md Shadman Aadeeb, et al. (2022). “A deep learning based multimodal interaction system for bed ridden and immobile hospital admitted patients: design, development and evaluation.” In: *BMC Health Services Research* 22.1, p. 803.

Islam, Muhammad Nazrul, Uzma Hasan, et al. (2022). “IoT-based serious gaming platform for improving cognitive skills of children with special needs.” In: *Journal of Educational Computing Research* 60.6, pp. 1588–1611.

Islam, Muhammad Nazrul and AKM Najmul Islam (2020). “A systematic review of the digital interventions for fighting COVID-19: the Bangladesh perspective.” In: *Ieee Access* 8, pp. 114078–114087.

Islam, Muhammad Nazrul, Iyolita Islam, et al. (2020). “A review on the mobile applications developed for COVID-19: an exploratory analysis.” In: *Ieee Access* 8, pp. 145601–145610.

Islam, Muhammad Nazrul, Md Karim, et al. (2020). “Investigating usability of mobile health applications in Bangladesh.” In: *BMC medical informatics and decision making* 20.1, pp. 1–13.

Islam, Muhammad Nazrul, Shahriar Rahman Khan, et al. (2021). “A mobile application for mental health care during covid-19 pandemic: Development and usability evaluation with system usability scale.” In: *Computational Intelligence in Information Systems: Proceedings of the Computational Intelligence in Information Systems Conference (CIIS 2020)*. Springer, pp. 33–42.

Islam, Muhammad Nazrul, Sumaiya Nuha Mustafina, et al. (2022). “Machine learning to predict pregnancy outcomes: a systematic review, synthesizing framework and future research agenda.” In: *BMC pregnancy and childbirth* 22.1, pp. 1–19.

Islam, Muhammad Nazrul, Ashratuz Zavin, et al. (2017). “GiveMed: a webportal for medicine distribution among poverty-stricken people.” In: *2017 IEEE Region 10 Humanitarian Technology Conference (R10-HTC)*. IEEE, pp. 294–299.

Jabbar, Meerja Akhil (2021). “Breast cancer data classification using ensemble machine learning.” In: *Engineering and Applied Science Research* 48.1, pp. 65–72.

Jamil, Avin S et al. (2015). “A case–control observational study of insulin resistance and metabolic syndrome among the four phenotypes of polycystic ovary syndrome based on Rotterdam criteria.” In: *Reproductive Health* 12.1, pp. 1–9.

Jia, Xibiao et al. (2020). “Endometrial cancer combined with polycystic ovary syndrome in 9 women under 40-years old: A case report.” In: *Biomedical Reports* 13.5, pp. 1–1.

Joham, Anju E and Helena J Teede (2022). “PCOS—a metabolic condition with health impacts on women and men.” In: *Nature Reviews Endocrinology* 18.4, pp. 197–198.

Kałużna, Małgorzata et al. (2020). “Effect of Central Obesity and Hyperandrogenism on Selected Inflammatory Markers in Patients with PCOS: A WHtR-Matched Case-Control Study.” In: *Journal of clinical medicine* 9.9, p. 3024.

Kaur, Harkawalpreet, Avleen Kaur Malhi, and Husanbir Singh Pannu (2020). “Machine learning ensemble for neurological disorders.” In: *Neural Computing and Applications*, pp. 1–18.

Kaur, Ranjeet, Amit Doegar, and Gaurav Kumar Upadhyaya (2022). “An Ensemble Learning Approach for Brain Tumor Classification Using MRI.” In: *Soft Computing: Theories and Applications*. Springer, pp. 645–656.

Kaushik, Shruti et al. (2020). “Ensemble of multi-headed machine learning architectures for time-series forecasting of healthcare expenditures.” In: *Applications of Machine Learning*. Springer, pp. 199–216.

Keerthi, S. Sathiya et al. (2001). “Improvements to Platt’s SMO algorithm for SVM classifier design.” In: *Neural computation* 13.3, pp. 637–649.

Khan, Nafiz Imtiaz et al. (2020). “Prediction of cesarean childbirth using ensemble machine learning methods.” In: *Proceedings of the 22nd international conference on information integration and web-based applications & services*, pp. 331–339.

Kharya, Shweta and Sunita Soni (2016). "Weighted naive bayes classifier: a predictive model for breast cancer detection." In: *International Journal of Computer Applications* 133.9, pp. 32–37.

Kiruthika, V and MM Ramya (2014). "Automatic segmentation of ovarian follicle using K-means clustering." In: *2014 fifth international conference on signal and image processing*. IEEE, pp. 137–141.

Kitchenham, Barbara (2004). "Procedures for performing systematic reviews." In: *Keele, UK, Keele University* 33.2004, pp. 1–26.

Kitchenham, Barbara and Stuart Charters (2007). "Guidelines for performing systematic literature reviews in software engineering." In.

Kondo, Masanari et al. (2019). "The impact of feature reduction techniques on defect prediction models." In: *Empirical Software Engineering* 24.4, pp. 1925–1963.

Kumar, P Suresh et al. (2021). "CatBoost ensemble approach for diabetes risk prediction at early stages." In: *2021 1st Odisha International Conference on Electrical Power Engineering, Communication and Computing Technology (ODICON)*. IEEE, pp. 1–6.

LaValley, Michael P (2008). "Logistic regression." In: *Circulation* 117.18, pp. 2395–2399.

Lawrence, Maryruth J et al. (2007). "Computer assisted detection of polycystic ovary morphology in ultrasound images." In: *Fourth Canadian Conference on Computer and Robot Vision (CRV'07)*. IEEE, pp. 105–112.

Lee, Tae-Hwy, Aman Ullah, and Ran Wang (2020). "Bootstrap aggregating and random forest." In: *Macroeconomic Forecasting in the Era of Big Data*. Springer, pp. 389–429.

Lu, Huijuan et al. (2019). "A hybrid ensemble algorithm combining AdaBoost and genetic algorithm for cancer classification with gene expression data." In: *IEEE/ACM transactions on computational biology and bioinformatics* 18.3, pp. 863–870.

Ma, Chaoqun et al. (2020). "Improved ORB algorithm using three-patch method and local gray difference." In: *Sensors* 20.4, p. 975.

Maini, Raman and Himanshu Aggarwal (2010). "A comprehensive review of image enhancement techniques." In: *arXiv preprint arXiv:1003.4053*.

Malini, NA and K Roy George (2018). “Evaluation of different ranges of LH: FSH ratios in polycystic ovarian syndrome (PCOS)–Clinical based case control study.” In: *General and comparative endocrinology* 260, pp. 51–57.

Mandal, Ardhendu, Debasmita Saha, and Manas Sarkar (2021). “Follicle Segmentation using K-means Clustering from Ultrasound Image of Ovary.” In: *Proceedings of International Conference on Frontiers in Computing and Systems*. Springer, pp. 545–553.

Maza, Sofiane and Mohamed Touahria (2019). “Feature selection for intrusion detection using new multi-objective estimation of distribution algorithms.” In: *Applied Intelligence* 49.12, pp. 4237–4257.

Meczekalski, Blazej et al. (2020). “The polycystic ovary syndrome and gynecological cancer risk.” In: *Gynecological Endocrinology* 36.4, pp. 289–293.

Meena, K, M Manimekalai, and S Rethinavalli (2015a). “Correlation of artificial neural network classification and NFRS attribute filtering algorithm for PCOS data.” In: *Int. J. Res. Eng. Technol* 4.3, pp. 519–524.

— (2015b). “Correlation of artificial neural network classification and NFRS attribute filtering algorithm for PCOS data.” In: *Int. J. Res. Eng. Technol* 4.3, pp. 519–524.

Mehrotra, Palak, Chandan Chakraborty, et al. (2011). “Automated ovarian follicle recognition for polycystic ovary syndrome.” In: *2011 International Conference on Image Information Processing*. IEEE, pp. 1–4.

Mehrotra, Palak, Jyotirmoy Chatterjee, et al. (2011). “Automated screening of polycystic ovary syndrome using machine learning techniques.” In: *2011 Annual IEEE India Conference*. IEEE, pp. 1–5.

Menger, Vincent, Floor Scheepers, and Marco Spruit (2018). “Comparing deep learning and classical machine learning approaches for predicting inpatient violence incidents from clinical text.” In: *Applied Sciences* 8.6, p. 981.

Mishra, Sushruta et al. (2021). “Thyroid disorder analysis using random forest classifier.” In: *Intelligent and cloud computing*. Springer, pp. 385–390.

Moran, Carmel M and Adrian JW Thomson (2020). “Preclinical ultrasound imaging—a review of techniques and imaging applications.” In: *Frontiers in Physics* 8, p. 124.

Mostafa, Sakib and Fang-Xiang Wu (2021). “Diagnosis of autism spectrum disorder with convolutional autoencoder and structural MRI images.” In: *Neural Engineering Techniques for Autism Spectrum Disorder*. Elsevier, pp. 23–38.

Munjal, Ashok, R Khandia, and B Gautam (2020a). “A machine learning approach for selection of polycystic ovarian syndrome (PCOS) attributes and comparing different classifier performance with the help of weka and pycaret.” In: *Int J Sci Res*, pp. 59–63.

— (2020b). “A machine learning approach for selection of polycystic ovarian syndrome (PCOS) attributes and comparing different classifier performance with the help of weka and pycaret.” In: *Int J Sci Res*, pp. 59–63.

Nabi, Nusrat et al. (2021). “Machine Learning Approach: Detecting Polycystic Ovary Syndrome & It’s Impact on Bangladeshi Women.” In: *2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT)*. IEEE, pp. 1–7.

Nadeem, Muhammad, Ayyaz Hussain, and Asim Munir (2019). “Fuzzy logic based computational model for speckle noise removal in ultrasound images.” In: *Multimedia Tools and Applications* 78.13, pp. 18531–18548.

Nanda, Sunpreet Kaur, Deepika Ghai, and Sagar Pande (2022). “VGG-16-Based Framework for Identification of Facemask Using Video Forensics.” In: *Proceedings of Data Analytics and Management*. Springer, pp. 673–685.

Nandini, Gangi Siva, AP Siva Kumar, and K Chidananda (2021). “Dropout technique for image classification based on extreme learning machine.” In: *Global Transitions Proceedings* 2.1, pp. 111–116.

Nandipati, Satish CR, CX Ying, and Khaw Khai Wah (2020a). “Polycystic Ovarian Syndrome (PCOS) classification and feature selection by machine learning techniques.” In: *Appl Math Comput Intell* 9, pp. 65–74.

— (2020b). “Polycystic Ovarian Syndrome (PCOS) classification and feature selection by machine learning techniques.” In: *Appl Math Comput Intell* 9, pp. 65–74.

Nilofer, NS et al. (2021a). “Follicles Classification To Detect Polycystic Ovary Syndrome Using Glcm And Novel Hybrid Machine Learning.” In: *Turkish Journal of Computer and Mathematics Education (TURCOMAT)* 12.7, pp. 1062–1073.

— (2021b). “Follicles Classification To Detect Polycystic Ovary Syndrome Using Glcm And Novel Hybrid Machine Learning.” In: *Turkish Journal of Computer and Mathematics Education (TURCOMAT)* 12.7, pp. 1062–1073.

Nusinovici, Simon et al. (2020). “Logistic regression was as good as machine learning for predicting major chronic diseases.” In: *Journal of clinical epidemiology* 122, pp. 56–69.

Octaviani, Theresia Lidya, Zuherman Rustam, and Titin Siswantining (2019). “Ovarian Cancer Classification using Bayesian Logistic Regression.” In: *IOP Conference Series: Materials Science and Engineering*. Vol. 546. 5. IOP Publishing, p. 052049.

Ogunleye, Adeola and Qing-Guo Wang (2019). “XGBoost model for chronic kidney disease diagnosis.” In: *IEEE/ACM transactions on computational biology and bioinformatics* 17.6, pp. 2131–2140.

Omuya, Erick Odhiambo, George Onyango Okeyo, and Michael Waema Kimwele (2021). “Feature selection for classification using principal component analysis and information gain.” In: *Expert Systems with Applications* 174, p. 114765.

Padmapriya, B and T Kesavamurthy (2015). “Diagnostic tool for PCOS classification.” In: *7th WACBE World Congress on Bioengineering 2015*. Springer, pp. 182–185.

Palomba, Stefano, Terhi T Piltonen, and Linda C Giudice (2021). “Endometrial function in women with polycystic ovary syndrome: a comprehensive review.” In: *Human reproduction update* 27.3, pp. 584–618.

Pang, Long et al. (2019). “A novel protein subcellular localization method with CNN-XGBoost model for Alzheimer’s disease.” In: *Frontiers in genetics* 9, p. 751.

Pathak, Hemita and Vrushali Kulkarni (2015). “Identification of ovarian mass through ultrasound images using machine learning techniques.” In: *2015 IEEE International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN)*. IEEE, pp. 137–140.

Prapty, Aroni Saha and Tanzim Tamanna Shitu (2020a). “An efficient decision tree establishment and performance analysis with different machine learning approaches on Polycystic Ovary Syndrome.” In: *2020 23rd International Conference on Computer and Information Technology (ICCIT)*. IEEE, pp. 1–5.

— (2020b). “An efficient decision tree establishment and performance analysis with different machine learning approaches on Polycystic Ovary Syndrome.” In: *2020 23rd International Conference on Computer and Information Technology (ICCIT)*. IEEE, pp. 1–5.

Prokhorenkova, Liudmila et al. (2018). “CatBoost: unbiased boosting with categorical features.” In: *Advances in neural information processing systems* 31.

Pulluparambil, Siji Jose and Subrahmanya Bhat (2021). “Medical Image Processing: Detection and Prediction of PCOS–A Systematic Literature Review.” In: *International Journal of Health Sciences and Pharmacy (IJHSP)* 5.2, pp. 80–98.

Purnama, Bedy et al. (2015a). “A classification of polycystic Ovary Syndrome based on follicle detection of ultrasound images.” In: *2015 3rd International Conference on Information and Communication Technology (ICoICT)*. IEEE, pp. 396–401.

— (2015b). “A classification of polycystic Ovary Syndrome based on follicle detection of ultrasound images.” In: *2015 3rd International Conference on Information and Communication Technology (ICoICT)*. IEEE, pp. 396–401.

Quinlan, J. Ross (1986). “Induction of decision trees.” In: *Machine learning* 1.1, pp. 81–106.

Rachana, B et al. (2021a). “Detection of polycystic ovarian syndrome using follicle recognition technique.” In: *Global Transitions Proceedings* 2.2, pp. 304–308.

— (2021b). “Detection of polycystic ovarian syndrome using follicle recognition technique.” In: *Global Transitions Proceedings* 2.2, pp. 304–308.

Rahman, Md Mahbubur et al. (2022a). “Hospital patients’ length of stay prediction: A federated learning approach.” In: *Journal of King Saud University-Computer and Information Sciences* 34.10, pp. 7874–7884.

Rahman, Md. Mahbubur et al. (2022b). “Hospital patients’ length of stay prediction: A federated learning approach.” In: *Journal of King Saud University - Computer and Information Sciences*. ISSN: 1319-1578. DOI: https://doi.org/10.1016/j.jksuci.2022.07.006. URL: https://www.sciencedirect.com/science/article/pii/S1319157822002336.

Raj, Ashika (2013). “Detection of cysts in ultrasonic images of ovary.” In: *International Journal of Science and Research (IJSR)* 2.8, pp. 185–189.

Rehman, Arshia et al. (2020). “A deep learning-based framework for automatic brain tumors classification using transfer learning.” In: *Circuits, Systems, and Signal Processing* 39.2, pp. 757–775.

Richhariya, Bharat et al. (2020). “Diagnosis of Alzheimer’s disease using universum support vector machine based recursive feature elimination (USVM-RFE).” In: *Biomedical Signal Processing and Control* 59, p. 101903.

Rodriguez Paris, Valentina et al. (2022). “The interplay between PCOS pathology and diet on gut microbiota in a mouse model.” In: *Gut Microbes* 14.1, p. 2085961.

Rok, Blagus and L Lusa (2013). “SMOTE for high-dimensional class-imbalanced data.” In: *BMC Bioinformatics* 14.1, pp. 106–121.

Rustam, Zuherman and Ni Putu Ayu Audia Ariantari (2018). “Comparison between support vector machine and fuzzy Kernel C-Means as classifiers for intrusion detection system using chi-square feature selection.” In: *AIP Conference Proceedings*. Vol. 2023. 1. AIP Publishing LLC, p. 020214.

Safiri, Saeid et al. (2022). “Prevalence, incidence and years lived with disability due to polycystic ovary syndrome in 204 countries and territories, 1990–2019.” In: *Human Reproduction*.

Sagi, Omer and Lior Rokach (2018). “Ensemble learning: A survey.” In: *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 8.4, e1249.

Al-Sarem, Mohammed et al. (2021). “Feature selection and classification using CatBoost method for improving the performance of predicting Parkinson’s disease.” In: *Advances on Smart and Soft Computing*. Springer, pp. 189–199.

Sarker, Iqbal H, ASM Kayes, and Paul Watters (2019). “Effectiveness analysis of machine learning classification models for predicting personalized context-aware smartphone usage.” In: *Journal of Big Data* 6.1, pp. 1–28.

Selçuk, Ayşe Adin (2019). “A guide for systematic reviews: PRISMA.” In: *Turkish archives of otorhinolaryngology* 57.1, p. 57.

Sengur, Abdulkadir (2012). “Support vector machine ensembles for intelligent diagnosis of valvular heart disease.” In: *Journal of medical systems* 36.4, pp. 2649–2655.

Setiawati, Eni, ABW Tjokorda, et al. (2015). “Particle swarm optimization on follicles segmentation to support PCOS detection.” In: *2015 3rd International Conference on Information and Communication Technology (ICoICT)*. IEEE, pp. 369–374.

Sharif, Adnan et al. (2018). “Exploring the opportunities and challenges of adopting augmented reality in education in a developing country.” In: *2018 IEEE 18th International Conference on Advanced Learning Technologies (ICALT)*. IEEE, pp. 364–366.

Sheikholeslami, Sina et al. (2021). “Autoablation: Automated parallel ablation studies for deep learning.” In: *Proceedings of the 1st Workshop on Machine Learning and Systems*, pp. 55–61.

Shorten, Connor and Taghi M Khoshgoftaar (2019). “A survey on image data augmentation for deep learning.” In: *Journal of Big Data* 6.1, pp. 1–48.

Shrivastav, Lokesh Kumar and Sunil Kumar Jha (2021). “A gradient boosting machine learning approach in modeling the impact of temperature and humidity on the transmission rate of COVID-19 in India.” In: *Applied Intelligence* 51.5, pp. 2727–2739.

Sitheswaran, Ranjitha and S Malarkhodi (2014). “An effective automated system in follicle identification for Polycystic Ovary Syndrome using ultrasound images.” In: *2014 International Conference on Electronics and Communication Systems (ICECS)*. IEEE, pp. 1–5.

sklearn.preprocessing.MinMaxScaler (2022). en. URL: `https://scikit-learn/stable/modules/generated/sklearn.preprocessing.MinMaxScaler.html` (visited on 04/19/2022).

Suha, Sayma Alam, M Akhtaruzzaman, and Tahsina Farah Sanam (2022). “A fuzzy model for predicting burn patients’ intravenous fluid resuscitation rate.” In: *Healthcare Analytics* 2, p. 100070.

Suha, Sayma Alam and Muhammad Nazrul Islam (2022). “An extended machine learning technique for polycystic ovary syndrome detection using ovary ultrasound image.” In: *Scientific Reports* 12.1, p. 17123.

Suha, Sayma Alam, Muhammad Nazrul Islam, et al. (2022). “Assessing usability of mobile applications developed for autistic users through heuristic and semiotic evaluation.” In: *Proceedings of International Joint Conference on Advances in Computational Intelligence: IJCACI 2021*. Springer, pp. 25–39.

Suha, Sayma Alam and Tahsina Farah Sanam (2022a). “A deep convolutional neural network-based approach for detecting burn severity from skin burn images.” In: *Machine Learning with Applications* 9, p. 100371.

— (2022b). “A Machine Learning Approach for Predicting Patient’s Length of Hospital Stay with Random Forest Regression.” In: *2022 IEEE Region 10 Symposium (TENSYMP)*. IEEE, pp. 1–6.

— (2022c). “Challenges and Prospects of Adopting Industry 4.0 and Assessing the Role of Intelligent Robotics.” In: *2022 IEEE World Conference on Applied Intelligence and Computing (AIC)*. IEEE, pp. 524–529.

Sumathi, M et al. (2021). “Study and detection of PCOS related diseases using CNN.” In: *IOP Conference Series: Materials Science and Engineering*. Vol. 1070. 1. IOP Publishing, p. 012062.

Suyanto, Suyanto et al. (2022). “A new nearest neighbor-based framework for diabetes detection.” In: *Expert Systems with Applications* 199, p. 116857.

Syapariyah, AN, A Saifudin, T Desyani, et al. (2020). “Feature selection techniques to choose the best features for Parkinsons disease predictions based on decision tree.” In: *Journal of Physics: Conference Series*. Vol. 1477. 3. IOP Publishing, p. 032008.

Tandon, Anushree et al. (2020). "Blockchain in healthcare: A systematic literature review, synthesizing framework and future research agenda." In: *Computers in Industry* 122, p. 103290.

Tanwani, Namrata (2020). "Detecting PCOS using machine learning." In: *Int J Modern Trends Eng Sci (IJMTES)* 7.1, pp. 1–20.

Tchito Tchapga, Christian et al. (2021). "Biomedical image classification in a big data architecture using machine learning algorithms." In: *Journal of Healthcare Engineering* 2021.

Tefagh, Ghazale et al. (2022). "Effect of vitamin E supplementation on cardiometabolic risk factors, inflammatory and oxidative markers and hormonal functions in PCOS (polycystic ovary syndrome): a systematic review and meta-analysis." In: *Scientific reports* 12.1, pp. 1–16.

Thapa, Niraj et al. (2020). "DeepSuccinylSite: a deep learning based approach for protein succinylation site prediction." In: *BMC bioinformatics* 21.3, pp. 1–10.

Thaseen, Ikram Sumaiya and Cherukuri Aswani Kumar (2017). "Intrusion detection model using fusion of chi-square feature selection and multi class SVM." In: *Journal of King Saud University-Computer and Information Sciences* 29.4, pp. 462–472.

Tiwari, Shamik et al. (2022). "SPOSDS: A Smart Polycystic Ovary Syndrome Diagnostic System Using Machine Learning." In: *Expert Systems with Applications*, p. 117592.

Usmani, Ambreen, Rehana Rehman, and Zehra Akhtar (2014). "ASSOCIATION OF BODY MASS INDEX AND DIETARY HABITS WITH OVARIAN AND UTERINE MORPHOLOGY WITH SUBFERTILE POLYCYSTIC OVARIAN SYNDROME." In: *JPMI: Journal of Postgraduate Medical Institute* 28.2.

Vedpathak, Shreyas (Dec. 2020). *PCOS dataset*. URL: `https://www.kaggle.com/shreyasvedpathak/pcos-dataset`.

Vikas, B, Y Radhika, and K Vineesha (2021). "Detection of Polycystic Ovarian Syndrome using Convolutional Neural Networks." In: *Int J Cur Res Rev| Vol* 13.06, p. 156.

Wang, Rui and Ben Willem J Mol (2017). "The Rotterdam criteria for polycystic ovary syndrome: evidence-based criteria?" In: *Human Reproduction* 32.2, pp. 261–264.

Wang, Xiaoyi and Rong Zhang (2022). "Clinical Value Analysis of Combined Vaginal Ultrasound, Magnetic Resonance Dispersion Weighted Imaging, and Multilayer Spiral CT in the Diagnosis of Endometrial Cancer Using Deep VGG-16 AdaBoost Hybrid Classifier." In: *Journal of Oncology* 2022.

Xia, Xiaoling, Cui Xu, and Bing Nan (2017). "Inception-v3 for flower classification." In: *2017 2nd International Conference on Image, Vision and Computing (ICIVC)*. IEEE, pp. 783–787.

Xiao, Ruyi et al. (2021). "Early diagnosis model of Alzheimer's disease based on sparse logistic regression." In: *Multimedia Tools and Applications* 80.3, pp. 3969–3980.

Xu, Shouzhi et al. (2022). "Rumor detection on social media using hierarchically aggregated feature via graph neural networks." In: *Applied Intelligence*, pp. 1–14.

Yadav, Neha et al. (2022). "HSV model-based segmentation driven facial acne detection using deep learning." In: *Expert Systems* 39.3, e12760.

Yaman, Emine and Abdulhamit Subasi (2019). "Comparison of bagging and boosting ensemble machine learning methods for automated EMG signal classification." In: *BioMed research international* 2019.

Yamashita, Rikiya et al. (2018). "Convolutional neural networks: an overview and application in radiology." In: *Insights into imaging* 9.4, pp. 611–629.

YILMAZ, Perihan Gülşah and Güzin ÖZMEN (2020). "Follicle Detection for Polycystic Ovary Syndrome by using Image Processing Methods." In: *International Journal of Applied Mathematics Electronics and Computers* 8.4, pp. 203–208.

Yoo, Seung Hoon et al. (2020). "Deep learning-based decision-tree classifier for COVID-19 diagnosis from chest X-ray imaging." In: *Frontiers in medicine* 7, p. 427.

Zaw, Hein Tun, Noppadol Maneerat, and Khin Yadanar Win (2019). "Brain tumor detection based on Naïve Bayes Classification." In: *2019 5th International Conference on Engineering, Applied Sciences and Technology (ICEAST)*. IEEE, pp. 1–4.

Zeng, Xiangyan, Yen-Wei Chen, and Caixia Tao (2009). "Feature selection using recursive feature elimination for handwritten digit recognition." In: *2009 Fifth International*

Conference on Intelligent Information Hiding and Multimedia Signal Processing. IEEE, pp. 1205–1208.

Zhang, Jiayong and Yanxi Liu (2004). “Cervical cancer detection using SVM based feature screening.” In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, pp. 873–880.

Zhang, Yiming, Ying Weng, and Jonathan Lund (2022). “Applications of Explainable Artificial Intelligence in Diagnosis and Surgery.” In: *Diagnostics* 12.2, p. 237.

Zhao, Huimin et al. (2019). “Fault diagnosis method based on principal component analysis and broad learning system.” In: *IEEE Access* 7, pp. 99263–99272.

Zhu, Youlian and Cheng Huang (2012). “An improved median filtering algorithm for image noise reduction.” In: *Physics Procedia* 25, pp. 609–616.

Zhuang, Fuzhen et al. (2020). “A comprehensive survey on transfer learning.” In: *Proceedings of the IEEE* 109.1, pp. 43–76.

Zounemat-Kermani, Mohammad et al. (2021). “Ensemble machine learning paradigms in hydrology: A review.” In: *Journal of Hydrology* 598, p. 126266.

APPENDIX A

DATA EXTRACTION TABLE OF SLR

Table A.1: Data profiling of the selected research articles

Cat.	Data Type	Data Source	Data Volume	Ref	Frq
1,2	Patient Records with symptoms & diagnosis results as features	Patient records from hospitals of India (Collected from open source Kaggle data repository Vedpathak 2020)	541 Patient records with 44 features	Danaei Mehr and Polat 2022b; Nandipati, Ying, and Wah 2020b; Munjal, Khandia, and Gautam 2020b; Prapty and Shitu 2020b; Tanwani 2020; Bharati, Podder, M. Mondal, et al. 2022a; Boomidevi and Usha 2021b; M. S. K. Inan, Ulfath, et al. 2021a; Hassan and Mirza 2020; Adla et al. 2021	10
		Online Survey data collected from women of Bangladesh	550 Patient records	Nabi et al. 2021	1
		Patients records from Ghosh Dastidar Institute for Fertility Research (GDIFR), Kolkata between March 2010 and April 2011	250 Patient records	Mehrotra, Chatterjee, et al. 2011	1
		Data from infertility treatment centers at Thrissur,India	540 Patient records, 23 attributes	Denny et al. 2019b	1
		Data from GEO dataset-NCBI	303 Patient records, 26 attributes	Meena, Manimekalai, and Rethinavalli 2015b	1
3,4,5,6	Ovarian Ultrasonography (USG) images from patients	Women's Health Imaging Research Laboratory, (WHIRL) in Saskatoon, Canada	70 images (33 PCOS & 37 normal ovaries)	Lawrence et al. 2007	1
		Nandhini Sri Diagnostic Center, India.	90 images: normal(30), cystic(25) and PCOS(35)	Gopalakrishnan and Iyapparaja 2021	1
		Medillab Diagnostics, Gulbarga,India	90 images	Hiremath and Tegnoor 2010	1
		Ovarian mass USG images collected from various hospitals	120 images	Pathak and Kulkarni 2015	1
		Sardjito Hospital Yogyakarta, Indonesia	100 images	Hartati, Musdholifah, et al. 2019	1
		Various anonymous online open source repositories and websites	Not Specified	Nilofer et al. 2021b; Dewi, Wisesty, et al. 2018; Padmapriya and Kesavamurthy 2015; Sumathi et al. 2021; B, Yalavarthi, and K 2021; Purnama et al. 2015b; Raj 2013; Rachana et al. 2021b; Kiruthika and Ramya 2014	9

Table A.2: ML algorithms used in different studies

Algorithms	Ref	Frq
Random Forest Classifier	Danaei Mehr and Polat 2022b; Nandipati, Ying, and Wah 2020b; Munjal, Khandia, and Gautam 2020b; Prapty and Shitu 2020b; Bharati, Podder, M. Mondal, et al. 2022a; Nabi et al. 2021; M. S. K. Inan, Ulfath, et al. 2021a; Denny et al. 2019b; Hassan and Mirza 2020; Gopalakrishnan and Iyapparaja 2021	10
Multi-Layer Perceptron(MLP)	Danaei Mehr and Polat 2022b; Bharati, Podder, M. Mondal, et al. 2022a; M. S. K. Inan, Ulfath, et al. 2021a	3
K-Nearest Neighbour(KNN)	Nandipati, Ying, and Wah 2020b; Prapty and Shitu 2020b; Tanwani 2020; Bharati, Podder, M. Mondal, et al. 2022a; Nabi et al. 2021; M. S. K. Inan, Ulfath, et al. 2021a; Denny et al. 2019b; Adla et al. 2021; Lawrence et al. 2007; Rachana et al. 2021b	10
Support Vector Machine(SVM)	Nandipati, Ying, and Wah 2020b; Prapty and Shitu 2020b; Nabi et al. 2021; M. S. K. Inan, Ulfath, et al. 2021a; Meena, Manimekalai, and Rethinavalli 2015b; Denny et al. 2019b; Hassan and Mirza 2020; Adla et al. 2021; Pathak and Kulkarni 2015; Lawrence et al. 2007; Gopalakrishnan and Iyapparaja 2021; Purnama et al. 2015b	12
Gaussian Naive Bayes	Nandipati, Ying, and Wah 2020b; Prapty and Shitu 2020b; Nabi et al. 2021; Mehrotra, Chatterjee, et al. 2011; Meena, Manimekalai, and Rethinavalli 2015b; Denny et al. 2019b; Hassan and Mirza 2020; Gopalakrishnan and Iyapparaja 2021	8
Decision Tree	Munjal, Khandia, and Gautam 2020b; Bharati, Podder, M. Mondal, et al. 2022a; Nabi et al. 2021; Meena, Manimekalai, and Rethinavalli 2015b	4
Logistic Regression	Tanwani 2020; Bharati, Podder, M. Mondal, et al. 2022a; Nabi et al. 2021; Mehrotra, Chatterjee, et al. 2011; Denny et al. 2019b; Hassan and Mirza 2020	6
Linear discriminant classifier	Lawrence et al. 2007; Gopalakrishnan and Iyapparaja 2021	2
Classification & Regression Trees (CART)	Denny et al. 2019b; Hassan and Mirza 2020	2
Light Gradient Boosting Model(LGBM)	Bharati, Podder, M. Mondal, et al. 2022a	1
Gradient Boosting(GB)	Bharati, Podder, M. Mondal, et al. 2022a; Nabi et al. 2021	2
Ensemble Extra Tree	Danaei Mehr and Polat 2022b; Munjal, Khandia, and Gautam 2020b	2
Adaptive Boosting(AdaBoost)	Danaei Mehr and Polat 2022b; Nandipati, Ying, and Wah 2020b; Bharati, Podder, M. Mondal, et al. 2022a; M. S. K. Inan, Ulfath, et al. 2021a	4
Categorical Boosting(CATBoost)	Bharati, Podder, M. Mondal, et al. 2022a	1
Extreme Gradient Boosting(XGBoost)	Nabi et al. 2021; M. S. K. Inan, Ulfath, et al. 2021a	2
Voting Hard Ensemble Classifier	Bharati, Podder, M. Mondal, et al. 2022a	1
Voting soft Ensemble Classifier	Bharati, Podder, M. Mondal, et al. 2022a	1
Artificial Neural Network(ANN)	Boomidevi and Usha 2021b; Meena, Manimekalai, and Rethinavalli 2015b; Nilofer et al. 2021b	3
Convolutional Neural Network (CNN)	Dewi, Wisesty, et al. 2018; Sumathi et al. 2021; B, Yalavarthi, and K 2021	3

Table A.3: Feature selection techniques used in different studies

Feature selection techniques	**Ref**
Sequential backward selection (SBS)	Danaei Mehr and Polat 2022b
Pearson method	Danaei Mehr and Polat 2022b
Random Forest embedded feat. selection	Danaei Mehr and Polat 2022b
Recursive Feature Elimination (RFE)	Nandipati, Ying, and Wah 2020b; Bharati, Podder, M. Mondal, et al. 2022a
Chi-Square feature selection	Nandipati, Ying, and Wah 2020b; M. S. K. Inan, Ulfath, et al. 2021a
Forward Selection & Backward Elimination Technique	Nandipati, Ying, and Wah 2020b
Genetic Algorithm	Munjal, Khandia, and Gautam 2020b
Decision Tree	Prapty and Shitu 2020b
Filter Method	Tanwani 2020
Univariate Feature Selection Method	Bharati, Podder, M. Mondal, et al. 2022a
Statistical Analysis(t-test)	Mehrotra, Chatterjee, et al. 2011
Analysis of variance (ANOVA) Test	M. S. K. Inan, Ulfath, et al. 2021a; Adla et al. 2021
Neural Fuzzy Rough Set (NFRS) & Artificial Neural Network (ANN)	Meena, Manimekalai, and Rethinavalli 2015b
Correlation based Feature Selection (CFS)	Meena, Manimekalai, and Rethinavalli 2015b
Principal Component Analysis(PCA)	Meena, Manimekalai, and Rethinavalli 2015b; Denny et al. 2019b

Table A.4: Performance metrics used for evaluating the proposed methodologies of the articles

Metrics	**Ref**	**Freq**
Accuracy	Danaei Mehr and Polat 2022b-Dewi, Wisesty, et al. 2018,Sumathi et al. 2021; B, Yalavarthi, and K 2021; Pathak and Kulkarni 2015; Lawrence et al. 2007; Purnama et al. 2015b; Rachana et al. 2021b	23
Sensitivity (or Recall)	Danaei Mehr and Polat 2022b-Nilofer et al. 2021b,Pathak and Kulkarni 2015; Lawrence et al. 2007	16
Specificity	Danaei Mehr and Polat 2022b; Mehrotra, Chatterjee, et al. 2011; Denny et al. 2019b; Pathak and Kulkarni 2015; Lawrence et al. 2007	5
Precision	Danaei Mehr and Polat 2022b-Nilofer et al. 2021b	16
F1-Score	Danaei Mehr and Polat 2022b; Munjal, Khandia, and Gautam 2020b; Prapty and Shitu 2020b; Tanwani 2020; Bharati, Podder, M. Mondal, et al. 2022a; Nabi et al. 2021; M. S. K. Inan, Ulfath, et al. 2021a; Denny et al. 2019b; Hassan and Mirza 2020; Nilofer et al. 2021b	10
Confusion Matrix	Munjal, Khandia, and Gautam 2020b; Bharati, Podder, M. Mondal, et al. 2022a; Nabi et al. 2021; M. S. K. Inan, Ulfath, et al. 2021a	4
Area Under the Curve(AUC) Score	Munjal, Khandia, and Gautam 2020b	1
Receiver Operating Characteristic(ROC) curve	Hassan and Mirza 2020	1
Positive Predictive Value (PPV)	Mehrotra, Chatterjee, et al. 2011	1
Loss	Hassan and Mirza 2020	1
Error Rate	Gopalakrishnan and Iyapparaja 2021; Hiremath and Tegnoor 2010	2
Comparative evaluation with manual process	Padmapriya and Kesavamurthy 2015; Lawrence et al. 2007; Gopalakrishnan and Iyapparaja 2021; Hiremath and Tegnoor 2010; Kiruthika and Ramya 2014	5