Advanced Diagnosis of Polycystic Ovarian Syndrome using Machine Learning and Multimodal Data Integration

Nethra Sai M[1], Sakthivel V[2], Prakash P[3], Vishnukumar K[4], Dugki Min[5]
School of Computer Science Engineering, Vellore Institute of Technology, Chennai, India[1, 2, 3]
Department of Computer Science of Engineering, KPR Institute of Engineering and Technology, Chennai, India[4]
Department of Computer Science of Engineering, Konkuk University, Seoul, South Korea[5]

***Abstract*—PCOS is a common endocrine disorder that impacts women in their reproductive years characterized by irregular menstrual cycles, hyperandrogenism, and polycystic ovaries. Polycystic Ovary Syndrome (PCOS) presents significant challenges in diagnosis due to its heterogeneous nature and varied clinical manifestations. This project aimed to develop a comprehensive system for PCOS detection, integrating ultrasound images and clinical data through advanced machine learning techniques, using Rotterdam criteria for diagnostic decisions. Feature extraction from ultrasound images was conducted using the ResNet-50 deep learning model, while clinical data underwent correlation-based feature selection. Three classification algorithms - Support Vector Machine (SVM), Random Forest and Logistic Regression - were used to categorize the extracted features from ultrasound images. The integration of image-based and clinical-based features was explored and evaluated to have better accuracy revealing the potential for enhancing PCOS diagnosis accuracy. The developed system holds promise for assisting doctors in PCOS diagnosis, offering a holistic approach that leverages both imaging and clinical information.**

***Keywords*—PCOS; ultrasound images; clinical data; feature extraction; classification; Rotterdam criteria**

I. Introduction

Polycystic Ovarian Syndrome (PCOS) is a prevalent hormonal condition affecting individuals in their reproductive years, with a global prevalence estimated to be between 8% and 13%. This multifaceted condition is characterized by a range of symptoms, including irregular menstrual cycles, elevated levels of androgens (hyperandrogenism), and the appearance of multiple cysts on the ovaries as seen on ultrasound. Despite its widespread impact, PCOS diagnosis remains challenging, often requiring a multidimensional assessment of clinical, biochemical, and imaging data. However, current diagnostic approaches often rely on the interpretation of disparate data sources, leading to variability and potential delays in diagnosis. Current systems often lack clarity on the criteria used for diagnosis, leading to inconsistencies and potential misdiagnoses.

In response to these challenges, this research aims to introduce an intelligent PCOS diagnostic system that leverages machine learning to integrate health records and ultrasound imaging. The integration of health records provides a rich source of clinical and biochemical data, encompassing information on menstrual patterns, hormonal levels, and other relevant patient history. Complementing this, the inclusion of ultrasound imaging allows for the examination of ovarian morphology, particularly the presence of multiple small follicles. By combining these diverse data modalities, the proposed system seeks to create a more comprehensive and accurate diagnostic framework.

The main objective of this research is to enhance PCOS diagnosis, facilitating early identification and intervention. By harnessing machine learning algorithms, this study aims to develop a model that can identify nuanced patterns in the data, thus improving the accuracy and effectiveness of PCOS diagnosis. The integration of health records and ultrasound imaging serves as a strategic foundation, recognizing the importance of a holistic approach to reproductive health. Additionally, this research employs the Rotterdam criteria, a widely accepted standard in PCOS diagnosis, to ensure that the decision-making process is grounded in established clinical guidelines. The outcome of this research offers potential not just for reproductive health but also for the wider realm of personalized medicine and data-driven healthcare innovations.

II. Background

PCOS is a multifaceted hormonal disorder impacting women during their reproductive years, noted for its varied clinical presentations and effects on metabolic health. The diagnosis of PCOS involves a multifaceted assessment of various criteria, reflecting the heterogeneity of the syndrome. The Rotterdam criteria, [1] frequently embraced in clinical practice, require the presence of a minimum of two out of three primary features for diagnosis: irregular menstrual cycles (oligo-anovulation), clinical or biochemical indicators of elevated androgens (hyperandrogenism), and the observation of multiple cysts on the ovaries during ultrasound examination.

- Oligo-anovulation refers to irregular menstrual cycles or the absence of menstruation. This criterion acknowledges the hormonal dysregulation that often underlies PCOS and is crucial for diagnosis.
- Hyperandrogenism manifests as elevated levels of androgens, leading to clinical symptoms such as acne, hirsutism (excessive hair growth), and male-pattern baldness. Biochemical evidence, including increased testosterone levels, supports this criterion.

- Ultrasound imaging plays a crucial role in the diagnosis of PCOS by providing a visual representation of the ovaries. The typical findings include the presence of 12 or more small follicles (2-9 mm in diameter) in each ovary and/or an enlarged ovarian volume.

Despite these established criteria, PCOS diagnosis remains challenging due to variations in symptom presentation and the potential overlap with other conditions. The reliance on clinical judgment, often subjective, underscores the need for objective and data-driven diagnostic approaches. This study aims to address this need by proposing an intelligent diagnostic system that leverages machine learning to integrate health records and ultrasound imaging, contributing to a more precise, timely, and personalized approach to PCOS diagnosis. The subsequent sections will delve into the methodology, data integration processes, and potential implications of this novel diagnostic framework.

III. Related Works

Many research papers focusing on PCOS detection predominantly center on the identification of cysts within ultrasound images through a variety of methodologies. The paper by M. Sumathi et al. [2] demonstrates the effectiveness of utilizing image processing techniques and classification algorithms such as DarkNet-19, AlexNet, SqueezeNet, and SVM for automated PCOS diagnosis. Gray Level Co-Occurrence Matrix was used for extracting features from the images and classification with DarkNet-19 achieved 99% accuracy, thus improving performance metrics. PCO follicle detection through preprocessing, feature extraction, and classification phases proposed by Bedy Purnama et al. utilizes techniques like Gabor wavelets for feature extraction, [3] it employs SVM-RBF Kernel for classification, achieving 82.55% accuracy for Dataset A and 78.81% for Dataset B, demonstrating its potential for enhancing PCOS diagnosis accuracy. Yinhui Deng et al. proposed object-growing algorithm [4] initially identifies multiple objects, likely follicles, with high probabilities from ultrasound images. It utilizes a cost map to differentiate the ovary from external regions and dynamically updates potential follicles based on their cost functions. This approach achieved an 89.4% recognition rate and a 7.45% misidentification rate on 31 actual PCOS ultrasound images, demonstrating superior performance compared to other methods. The method by Sharvari S Deshpande et al. uses ovarian ultrasound image processing, feature extraction, segmentation, and classification through Support Vector Machine (SVM), achieving a high accuracy of 95%. [5] The research employs preprocessing techniques, including contrast enhancement and filtering, on ovarian ultrasound images. Feature extraction involves Multiscale morphological approach and Top-hat transform, while segmentation uses Canny edge detection. The approach proposed by SaymaAlma Suham et al. involves employing a CNN with transfer learning for feature extraction and a stacking ensemble machine learning model with XGBoost [6] as the meta-learner for classification. The proposed technique achieves accuracy improvement, reaching 99.89%, with reduced execution time compared to existing machine learning methods. The optimal performance is achieved by integrating the "VGGNet16" pre-trained model with CNN for feature extraction and utilizing "XGBoost" as the meta-learner for classification. An innovative machine learning approach was proposed by Pradeep Bedi et al., the Attention Residual UNet (AResUNet) Model, [7] for detecting PCOS. The model incorporates adaptive bilateral filter-based image pre-processing with attention-guided residual UNet allowing it to effectively handle both 2D and multi-modal images. The results demonstrate that the AResUNet Model achieves high accuracy of 98%. The method proposed by Asma' Amirah Nazarudin et al. combines Otsu's thresholding with the Chan–Vese method [8] to create a binary mask and define follicle boundaries. Compared to the classical Chan–Vese method, the proposed approach demonstrated superior performance with an average sensitivity of 0.74, which was significantly higher than the sensitivity of 0.54 for the classical Chan–Vese method.

Certain papers utilize clinical data, comprising information from manually recorded ultrasound images by radiologists, alongside other parameters crucial for PCOS detection. A balanced dataset was achieved Ejay Nsugbe by using synthetic sample generation software to mitigate bias in training prediction models. [9] Ten machine learning models were explored, revealing high-order SVM with a nonlinear decision boundary as the optimal classifier demonstrating superior performance. The research by Satish C. R Nandipati et al. aims to identify the most effective classification model and significant features for predicting PCOS, utilizing Python-Scikit Learn and RapidMiner tools. The results in [10] indicate that Random Forest achieves the highest accuracy (93.12%, RapidMiner) with the complete dataset, KNN and SVM exhibit similar accuracy (90.83%, RapidMiner) with 10 selected features. ML is employed to construct a stacking ensemble model by Hela Elmannai et al. combining LR, RF, DT, NB, SVM, KNN, Xgboost, and Adaboost [11] at the base learner level, with RF at the meta-learner level, aiming to enhance single ML performance. The resulting Stacking ML, particularly with REF feature selection, achieved notable performance recording high accuracy 98.87%. The application of ensemble classifiers, including Ensemble Random Forest, Extra Tree, Adaptive Boosting (AdaBoost), and Multi-Layer Perceptron (MLP) for diagnosing PCOS is explored [12] by Homay Danaei Mehr et al. Subrato Bharati et al.'s study compares classifiers using holdout and cross-validation methods. Their results show that ensemble Random Forest, with feature subset selection [13], achieves the highest accuracy of 98.89% and sensitivity of 100%. RFLR demonstrates the highest testing accuracy of 91.01% and a recall value of 90% when using 40-fold cross-validation on these 10 most important features. A novel feature selection method proposed by Shazia Nasim et al., optimized chi-squared (CS-PCOS) [14] to select required features for detecting PCOS. Among ten hyper-parameterized machine learning models, Gaussian Naive Bayes (GNB) excelled, achieving 100%. Employing MATLAB and a dataset from Kaggle, paper by Dana Hdain et al. [15] utilized seven classifiers, with Linear Discriminant exhibiting the highest accuracy and K-Nearest Neighbor showing the best sensitivity for detection of PCOS. The paper by Manjunathan Alagarsamy et al. employs preprocessing techniques, such as a heat map for feature correlation, and utilizes Support Vector Machine, K-Nearest Neighbors, Naive Bayes, and a Hybrid Algorithm [16] for classification. The proposed approach

demonstrates superior performance compared to other methods achieving high accuracies (97% for Ensemble, 95% for SVM, and 93% for Naive Bayes) in identifying PCOS-affected ovaries. The research by Amsy Denny et al. utilizes machine learning techniques, including Logistic regression, Naïve Bayes, and Random Forest Classifier (RFC), and identifies RFC [17] as the most accurate method with 89.02% accuracy. The machine learning models were implemented in Spyder Python IDE, and the system employed RFC after optimizing features with Principal Component Analysis (PCA). The article by Sayma Alam Suha et al. introduces a unique stacked ensemble approach [18]which combines weak traditional ML classifiers and boosting or bagging models, achieving 95.7% accuracy. It also explores feature selection techniques, with PCA identifying the top 25 features for effective forecasting.

The paper by Jay Jojo Cheng et al. aimed to develop ML algorithms for classifying polycystic ovary morphology in pelvic ultrasounds, [19] utilizing electronic medical records. Pelvic ultrasound reports from 39,093 patients were analyzed. The classifiers Gradient Boosted Tree text classifier and rule-based text classifier achieved high accuracy, with rates of 97.6% and 96.1% on the evaluation set of 1000 ultrasound reports. The paper by Victor Castro et al. aims to enhance the accuracy of identifying PCOS subjects [20] by utilizing electronic medical records text and data, compared to the conventional use of International Classification of Diseases 9 codes. A natural language processing approach was employed to identify PCOS subjects in electronic medical records, and an algorithm was developed using 32 terms to categorize definite PCOS cases based on Rotterdam criteria. The algorithm demonstrated a 64% confirmation rate for definite PCOS cases with a 9% false positive rate, comparable to the 66% confirmation rate using ICD-9 codes with an 8.5% false positive rate.

The work by Alamoudi et al. utilized [21] fine-tuned Inception architecture to classify ultrasound images, achieving 84.81% accuracy. Additionally, a study combining image and clinical features through deep learning showed promising results, with joint fusion type I outperforming, highlighting the significance of clinical data in PCOS diagnosis.

The above studies utilizes either ultrasound or clinical data to predict PCOS and lacks a standardized prediction basis. In contrast, this research employs a hybrid dataset and utilizes the Rotterdam criteria, endorsed by the NIH and widely used by medical professionals for diagnosing PCOS. This approach integrates both clinical and ultrasound data, providing a more comprehensive and validated method for PCOS prediction.

IV. Proposed System

The proposed PCOS detection system strategically combines deep learning and traditional machine learning methodologies for a thorough examination of ultrasound images and clinical data as shown in Fig. 1 below. The initial phase involves the application of a ResNet50 deep learning model to meticulously extract intricate features from ultrasound images leveraging the model's prowess in intricate pattern recognition.

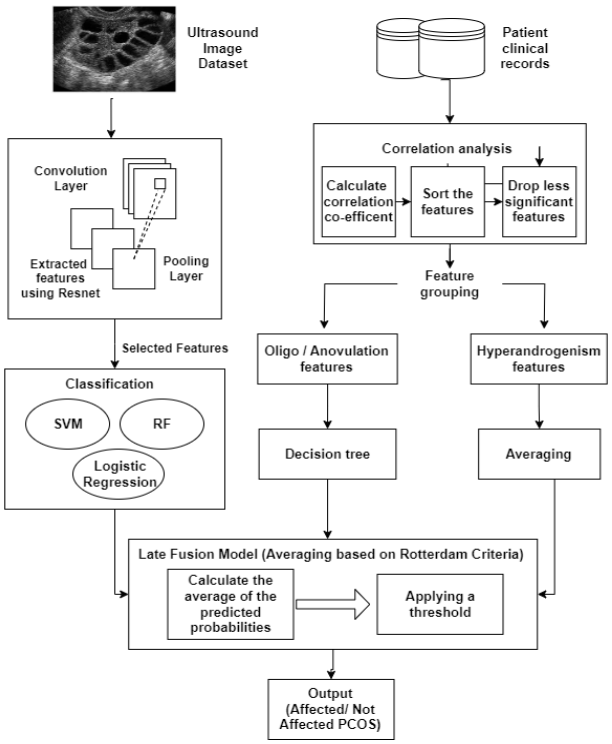

Fig. 1. System architecture to detect PCOS through ML and MMI.

These extracted features are pivotal for subsequent image classification, a crucial step in discerning whether the ultrasound images exhibit the characteristic traits of polycystic ovaries. Various classification models such as Support Vector Machines, Random Forest, and Logistic Regression are compared to find the best suited model.

Simultaneously, the clinical dataset undergoes meticulous curation through the selection of significant features, a process driven by correlation coefficients. This ensures that the subsequent analyses are streamlined, focusing on the most relevant clinical indicators. The chosen features are then segregated into two distinct categories: Oligo/anovulation features and hyperandrogenism features. For anovulation or oligo features, a decision tree model is deployed, bringing a level of interpretability to the assessment of irregular ovulation patterns. Concurrently, hyperandrogenism features undergo a nuanced analysis, with an averaging mechanism applied to gauge the extent of androgen excess.

The culmination of these diverse analyses leads to three distinct outputs, each providing a critical aspect of the PCOS diagnosis. The late fusion model, a mechanism that assigns weighted averages to the outputs, orchestrates the integration of these results. This step is pivotal, allowing for a nuanced combination of image-based evidence and clinically derived insights. The result of this intricate fusion process is the ultimate determination of whether the patient is affected with PCOS.

A. Ultrasound Feature Extraction

In the feature extraction process using the ResNet-50 model for ultrasound images, the aim is to capture and distill complex hierarchical features that are essential for subsequent

classification tasks. The ResNet-50 model, pre-trained on the ImageNet dataset, is employed as a feature extractor due to its proficiency in discerning intricate patterns and representations within images. The model, consisting of 50 layers, is well-suited for this task as it possesses a deep architecture that allows for the extraction of high-level features. The dataset consists of 3200 ultrasound images for training and 1468 images for testing. Each ultrasound image is loaded and resized to a standard size of 224x224 pixels. It is then converted into a numerical array and preprocessed to meet the ResNet-50 model's specifications. The pre-trained ResNet-50 model is then utilized to predict the features present in the image. ResNet-50 operates by passing the input image through a series of convolutional layers, pooling layers, and activation functions. The convolutional layers within ResNet-50 are designed to learn hierarchical features of increasing complexity. Lower layers of the network learn simple patterns like edges and textures, while deeper layers gradually extract more complex and distinctive features relevant to the task.

In each convolutional layer, the operation can be represented mathematically as a convolution operation followed by a non-linear activation function. Let us denote the output of the convolutional layer as H_l where l is the layer index. The mathematical operation for a single convolutional layer can be represented in Eq. (1):

$$H_l = \sigma(W_l * H_{l-1} + b_l) \quad (1)$$

where, W_l is the set of learnable weights (filters for layer l, H_{l-1} is the input feature map from the previous layer, b_l is the bias term, σ is the activation function (commonly ReLU). Throughout the network, the Rectified Linear Unit (ReLU) activation function is frequently employed to add non-linearity following each convolutional layer. ReLU function is defined as shown in Eq. (2).

$$f(x) = \max(0, x) \quad (2)$$

ResNet-50 makes use of residual blocks, which alleviate the vanishing gradient issue and facilitate the training of deep networks by introducing skip connections. Each residual block within the network contains a shortcut connection, allowing the input to bypass certain layers and directly propagate to deeper layers. Residual block is mathematically represented as shown in Eq. (3)

$$H_l = F(H_{l-1}, \{W_{l,i}\}) + H_{l-1} \quad (3)$$

where F represents the residual function $\{W_{l,i}\}$ denotes the set of learnable weights specific to the residual block and H_{l-1} is the input to the block. Global average pooling is used to transform the spatial data into a vector representation at the end of the convolutional layers. It is indicated mathematically through Eq. (4).

$$v = \frac{1}{N}\sum_{i=1}^{N} H_i \quad (4)$$

where v is the vector representation of the image features, N is the number of elements in the feature map and H_i represents the individual elements of feature map. For ultrasound images, these features might include distinctive patterns related to ovarian structures, cysts, or other relevant characteristics indicative of polycystic ovaries. The resulting feature vector is flattened to create a one-dimensional array, capturing the essence of the image's intricate characteristics. The features are then saved in a NumPy (.npy) file format, creating a reusable and compact representation that can be easily utilized in subsequent stages of the PCOS detection system, such as model training and classification.

B. Ultrasound Image Classification

The objective extends beyond a binary determination of the presence or absence of polycystic ovaries. Specifically, the classification system aims to discern nuances within the ultrasound images, including the identification of minimal and small-sized cysts, which are categorized as healthy and unhealthy ovaries are denoted with a greater number of cysts which are bigger in size. Three distinct classification models, namely Support Vector Machines (SVM), Random Forest, and Logistic Regression, are employed to discern patterns within these features and make predictions based on the labeled training data.

SVM, Random Forest, and Logistic Regression all analyze ResNet-50 features extracted from ultrasound images. SVM finds an optimal separation line, Random Forest combines multiple decision trees for complex patterns, and Logistic Regression estimates the likelihood of PCOS. Comparing their performance helps identify the best model for accurate PCOS classification, improving your overall detection framework.

C. Clinical Feature Selection and Analysis

A systematic and data-driven approach is employed to distill relevant information from a dataset comprising 39 features. The initial step involves a careful consideration of feature correlation coefficients, allowing for the identification and subsequent removal of features with less impact on the overall analysis. Correlation coefficients are statistical measures that quantify the direction and strength of the relationship between two variables. This curation process is crucial as it optimizes the dataset, focusing on attributes that exhibit stronger relationships with the outcomes of interest. The Pearson correlation coefficient is the default approach used by Python's corr() method to determine the correlation between columns in a Data Frame containing numerical data. The linear relationship between two continuous variables is measured by the Pearson correlation coefficient. Given two variables X and Y, with observations (x_i, y_i) for I =1, 2, 3, ……n, the Pearson coefficient r, is calculated as shown in Eq (5).

$$r = \frac{\Sigma(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\Sigma(x_i-\bar{x})^2}\sqrt{\Sigma(y_i-\bar{y})^2}} \quad (5)$$

The features that survive this correlation-based filtering are then categorized into two distinct sets: those features contributing to oligo/anovulation and those features associated with hyperandrogenism.

1) Ovarian dysfunction: The irregular menstrual cycles that characterize oligo/anovulation can be attributed to either infrequent periods (oligomenorrhea) or the total lack of ovulation and menstruation (anovulatory cycles). The main feature is the irregularity in cycle duration, which differs from the normal range of 21 to 35 days for menstrual periods. From the selected features after correlation analysis, features related

to menstrual cycle like cycle length(days), irregular cycles etc are grouped together into this category. For the set of features contributing to oligo/anovulation, a decision tree model is utilized. Decision trees are particularly effective in scenarios where complex decision-making processes depend on multiple factors. In the context of PCOS detection system, the decision tree scrutinizes the features relevant to oligo/anovulation, aiming to create a clear and interpretable decision path. The flow of decisions that decide if the person has ovarian dysfunction is shown in Fig. 2. This model facilitates the determination of whether an individual has a likelihood of PCOS based on the presence of oligo or anovulation, providing valuable insights into menstrual irregularities that are characteristic of the syndrome.

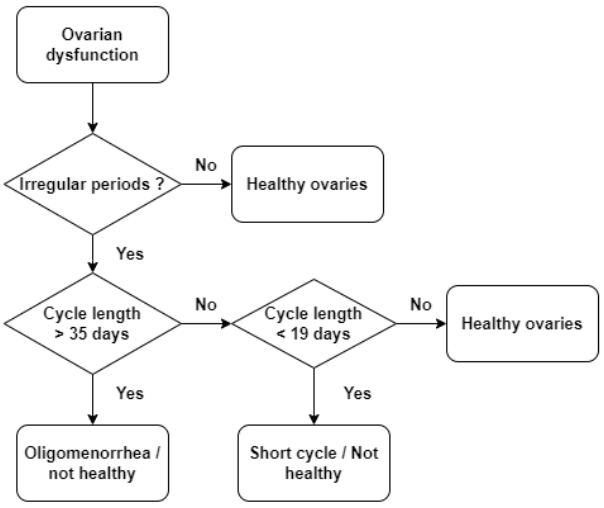

Fig. 2. Flow of decisions for features grouped into oligo/anovulation.

2) Hyperandrogenism features: Simultaneously, the features associated with hyperandrogenism undergo a distinct analytical process determined by averaging out the contributing factors. Features related to hyperandrogenism is grouped into this category after correlation analysis like skin darkening, pimples, hair loss etc. This approach acknowledges the multifactorial nature of hyperandrogenism, where diverse clinical indicators collectively contribute to the assessment of androgen excess. Averaging allows for a comprehensive and nuanced evaluation, providing a more accurate representation of the overall hyperandrogenic status. The integration of these clinical insights with the earlier image-based classification results in the final determination of PCOS presence, contributing to a holistic and robust diagnostic framework.

D. Data Integration and Final Prediction

In the data integration and final prediction stage of your PCOS detection system, the diverse outputs obtained from the ultrasound image SVM classifier, the oligo/anovulation decision tree, and the hyperandrogenism assessment are harmonized to yield a comprehensive and conclusive diagnosis. The distinct sources of information are treated as complementary dimensions contributing to the overall understanding of PCOS.

The SVM classifier, trained on the extracted features from ultrasound images, provides a binary classification output, discerning between individuals with and without polycystic ovaries. This classification serves as a foundational element in the final prediction, capturing crucial insights derived from the imaging data. Simultaneously, the decision tree model for oligo/anovulation offers a nuanced perspective on menstrual irregularities, helping identify individuals who exhibit characteristics associated with PCOS. This element enriches the diagnostic process by incorporating clinical indicators related to reproductive health aligning with the multifaceted nature of the syndrome. The hyperandrogenism assessment, determined through averaging clinical features, contributes a continuous and graded evaluation of androgen excess. This dimension acknowledges the spectrum of androgenic manifestations, offering a more refined understanding of the hormonal aspects of PCOS. The integration of these three outputs is orchestrated through a late fusion model, specifically a weighted average. The late fusion model allows for the consideration of the diverse nature of the inputs, assigning appropriate weights to each source based on their relative significance in the diagnostic process. But in this scenario as per Rotterdam criteria, equal weights have been assigned to the different inputs. Upon applying the weighted average, a continuous score is generated, reflecting the amalgamated insights from the image-based classification, reproductive health assessment, and hormonal evaluation. To finalize the prediction, a threshold is established, delineating the boundary between a positive and negative diagnosis for PCOS. This threshold serves as a decision criterion, guiding the system to categorize individuals based on the combined evidence from ultrasound images and clinical data.

The different phenotypes of PCOS are based on the presence or absence of these three features.

- Type A: This phenotype is the most prevalent and is distinguished by the presence of all three features: excess androgen levels, ovarian dysfunction, and polycystic ovarian morphology.
- Type B: This phenotype is defined by elevated androgen levels and ovarian dysfunction, although the ovaries do not exhibit the typical morphology associated with PCOS.
- Type C: This phenotype is marked by elevated androgen levels and the presence of polycystic ovarian morphology, despite normal ovarian function.
- Type D: This phenotype is distinguished by ovarian dysfunction and polycystic ovarian morphology, without the presence of elevated androgen levels.

These are just four of the many possible phenotypes of PCOS. The condition can vary greatly from woman to woman, and some women may have symptoms that do not fit neatly into any one category. The number of people affected with each phenotype is calculated to understand the diversity of the syndrome.

V. Results and Discussion

ResNet-50 extracted features from both training and testing datasets, yielding arrays of 100352 dimensions (224,224,2). Subsequently, three classification models – Support Vector Machine (SVM), Random Forest and Logistic Regression - were utilized to categorize the extracted features. Evaluation metrics were computed for each model to gauge their effectiveness. For SVM, the results demonstrated high accuracy (0.99), precision (0.99), and F1-score (0.98), along with a respectable AUC-ROC value of 0.98. Random Forest exhibited slightly lower accuracy (0.95) and AUC-ROC (0.89), but still showed strong precision (0.97) and F1-score (0.87). Logistic Regression performed consistently well across metrics, with accuracy at 0.97, precision at 0.92, recall at 0.92, F1-score at 0.92, and AUC-ROC at 0.95.

TABLE I. Comparison of Evaluation Metrics for Different Classification Models

Evaluation Metrics	SVM	Random Forest	Logistic Regression
Accuracy	99	95	97
Precision	99	97	92
Recall	96	79	92
F1 - score	98	87	92

Furthermore, Receiver Operating Characteristic (ROC) curves were plotted for all three models, providing visual insights into their performance. The ROC curve illustrates the true positive rate (TPR) on the y-axis against the false positive rate (FPR) on the x-axis, as depicted in Fig. 3. The area under the ROC curve (AUC) is a metric indicating the model's ability to differentiate between positive and negative cases. A perfect model would have an AUC of 1. In the provided ROC curve, the SVM model has the highest AUC (0.98), followed by the Logistic Regression model (0.95) and the Random Forest model (0.89). This suggests that the SVM model is the most effective at distinguishing between positive and negative cases in this scenario. The summarized results are mentioned in Table I.

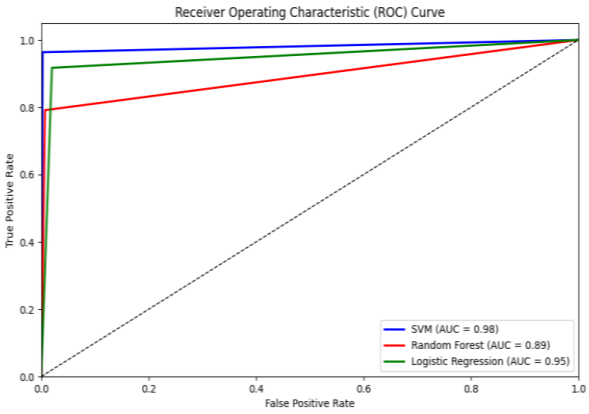

Fig. 3. ROC curve for three classification models.

The analysis also included SHAP interpretability plots for each model, enhancing the understanding of feature importance and contribution to classification decisions. Fig. 4, Fig. 5 and Fig. 6 displays the mean magnitude of SHAP values for each feature, showing both the direction and strength of the impact. Each dot represents a specific instance, and the color indicates the feature value (red for high values, blue for low values). Features are ordered by their importance based on the mean absolute SHAP values across all instances. The most influential features are located at the top. The horizontal position of each dot reflects the impact of the corresponding feature on the model's prediction for a specific instance. Dots positioned to the left contribute negatively, while those on the right contribute positively. The summary plot helps you understand the contribution of each feature to the model's predictions across different instances in your dataset.

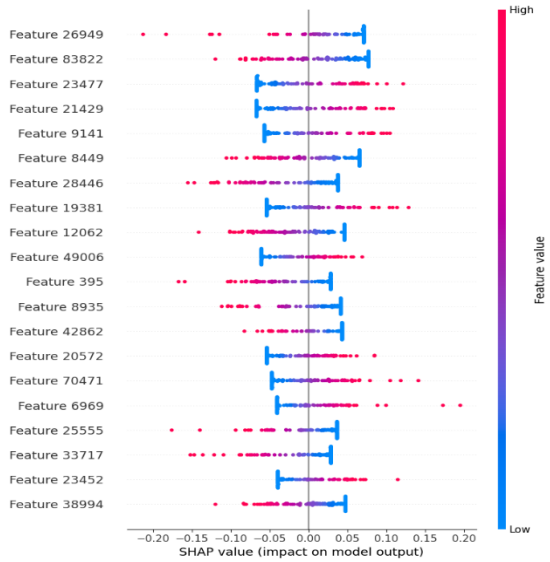

Fig. 4. SHAP interpretability of SVM classifier.

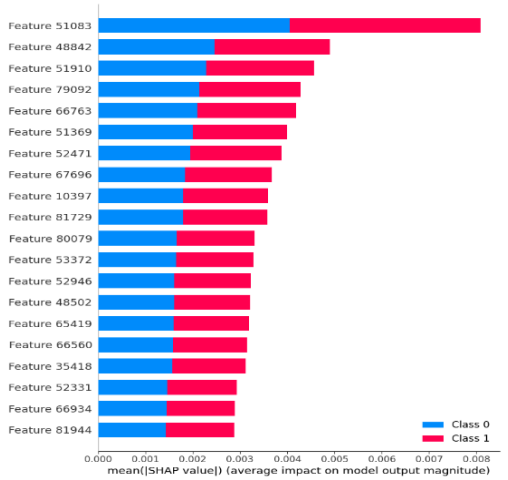

Fig. 5. SHAP interpretability of Random Forest classifier.

In the feature extraction process for clinical data, correlation coefficients were computed for various features, including Age (yrs), Weight (Kg), Height (Cm), BMI, Blood Group, Pulse rate (bpm), and others, which is plotted in x-axis. The y-axis shows the correlation coefficient, which is a measure of how strong the relationship is between a particular feature and PCOS. It can range from -1 to 1, where -1 indicates a perfect negative

correlation, 0 indicates no correlation, and 1 indicates a perfect positive correlation. Features with correlation coefficients above 17% were considered significant and selected for further analysis. The selected features include Weight (Kg), BMI, Cycle (R/I), Cycle length (days), Weight gain (Y/N), Skin darkening (Y/N), hair growth (Y/N), Hair loss (Y/N), Pimples (Y/N) and Fast food (Y/N). These features exhibit substantial correlations with the target variable or are deemed clinically relevant for PCOS diagnosis, thereby enhancing the effectiveness of subsequent analysis and modeling efforts.

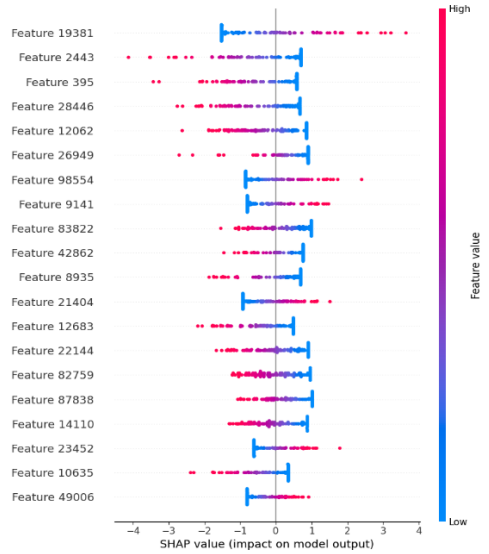

Fig. 6. SHAP interpretability of Logistic regression classifier.

The selected features are grouped into two categories, features contributing to hyperandrogenism and oligo/anovulation as shown below in Fig. 7. A decision tree helps assess irregular periods and Rotterdam criteria. Irregular periods with abnormal cycle length (less than 19 or more than 35 days) suggest oligomenorrhea, while normal cycle length (19-35 days) indicates healthy ovaries. Regular periods also suggest healthy ovaries. In the hyperandrogenism assessment, the features skin darkening, pimples, hair growth, hair loss, and fast-food consumption are collectively evaluated. Averaging these features provides a holistic perspective on the presence of hyperandrogenism, a common manifestation of PCOS. By combining these indicators, the analysis aims to capture the overall pattern of hyperandrogenic symptoms, offering a simplified yet comprehensive approach to assessing this aspect of PCOS.

The late fusion model integrates outputs from three distinct features- polycystic ovaries, hyperandrogenism, and oligo/anovulation to provide a unified assessment of PCOS diagnosis. Utilizing weighted averaging, each feature is assigned equal importance based on Rotterdam criteria, ensuring a balanced consideration of all contributing factors. If more than two inputs indicate the presence of PCOS, the individual is classified as having the condition, thereby delineating four distinct phenotypes. This approach facilitates a comprehensive evaluation of PCOS status, accounting for the multifaceted nature of the syndrome and enabling tailored treatment strategies based on identified phenotypes.

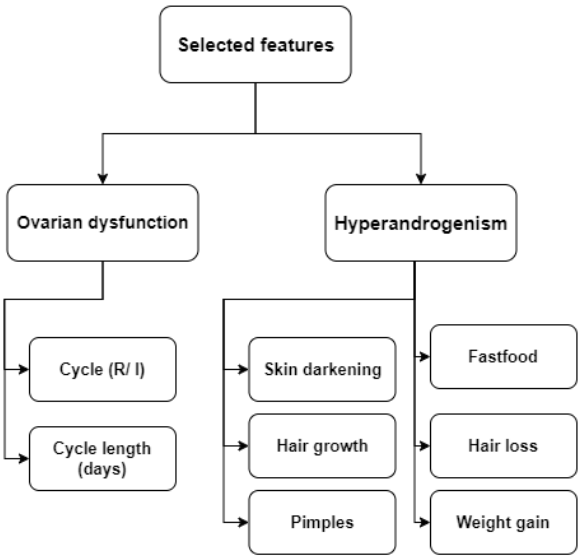

Fig. 7. Correlation analysis of clinical features.

The late fusion model uses weighted averaging and determines whether the patient is affected by PCOS or not. Fig. 8 shows the result of the fusion model in which out of the 1468 patients, 464 patients have PCOS whereas 1004 patients are not affected with PCOS.

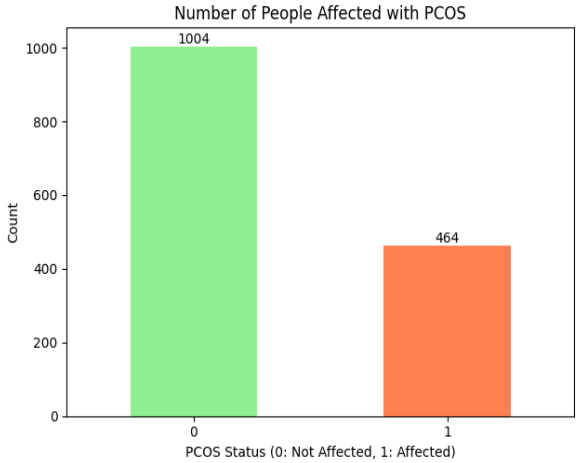

Fig. 8. Number of people affected vs. number of people not affected in total of 1468 patients.

The bar graph in Fig. 9 shows the distribution of people with and without polycystic ovary syndrome (PCOS) across three features that are commonly used to diagnose the condition according to Rotterdam criteria. There is fair share of affected and not affected for each feature which shows that each of the feature is important to accurately diagnose if the person has PCOS or not. It is important to note that not all women with PCOS will exhibit all these characteristics. Diagnosis typically involves a combination of symptoms, physical signs, and diagnostic tests such as blood tests and ultrasound imaging.

The affected patients (464 people) can be further categorized into 4 phenotypes to see the different possible combinations of features. The presence of all three features in a

patient gives Type A and the presence of two out of three features gives the rest of the phenotypes. (Type B, Type C, Type D). The number of people in different phenotypes are calculated and shown in Fig. 10. Type B is the phenotype with highest number of patients which shows that hyperandrogenism and ovarian dysfunction is equally important as polycystic ovaries for diagnosis of PCOS.

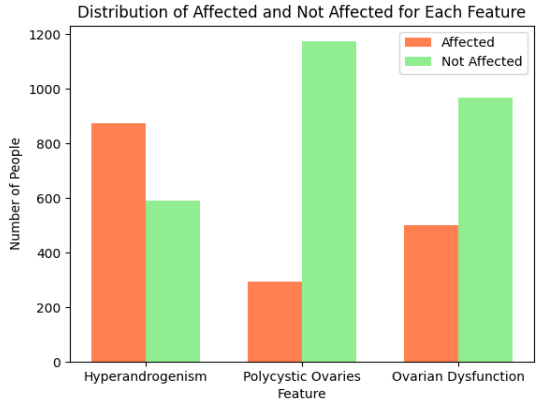

Fig. 9. Distribution of patients over three features mentioned as per Rotterdam criteria.

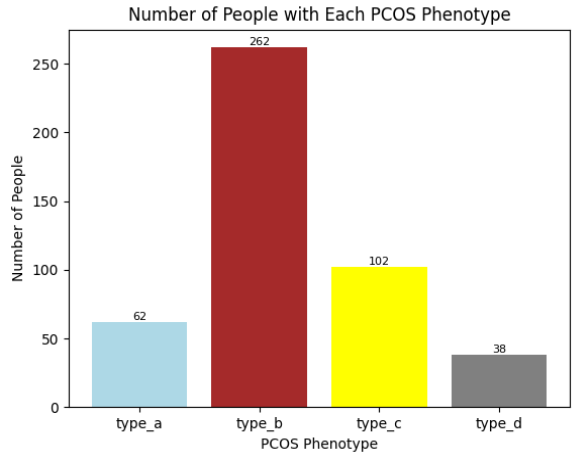

Fig. 10. Distribution of PCOS affected patients over different phenotypes.

The heatmap in Fig 11 appears to show the feature correlations between three features related to PCOS diagnosis: ovarian polycystic status, ovarian dysfunction, and hyperandrogenism. Each square in the heatmap illustrates the correlation coefficient between two features, varying from -1 (perfect negative correlation) to 1 (perfect positive correlation), with 0 indicating no correlation. The color intensity indicates the strength of the correlation, with darker shades denoting stronger correlations.

- Ovarian polycystic status: This feature has a strong positive correlation with both ovarian dysfunction (0.73) and hyperandrogenism (0.6). This suggests that patients with polycystic ovaries are more likely to also have ovarian dysfunction and hyperandrogenism, which are all characteristics of PCOS.
- Ovarian dysfunction: This feature has a moderate positive correlation with hyperandrogenism (0.4). This indicates that there is a positive association between these two features, but the relationship is not as strong as the one between ovarian polycystic status and the other two features.
- Hyperandrogenism: This feature has a weak positive correlation with ovarian polycystic status (0.6) and a moderate positive correlation with ovarian dysfunction (0.4). This suggests that hyperandrogenism is associated with both PCOS risk factors, but the strength of the association varies.

Overall, the heatmap confirms that there are positive correlations between all three features, which is consistent with the established risk factors for PCOS.

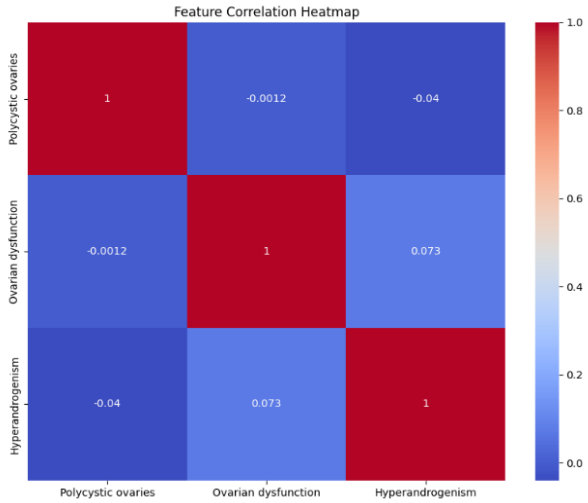

Fig. 11. Heatmap of feature correlation.

A. Validation of Results

The proposed system is evaluated on different evaluation metrics. The predictions are evaluated against the patient's actual health which is validated by a medical expert. The proposed system achieved promising performance metrics in the evaluation. Accuracy is the ratio of correctly classified instances to the total number of instances. It serves as a metric for assessing the overall performance of a model.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (6)$$

In Eq. (6), Eq. (7), Eq. (8), Eq. (9) TP stands for True Positive, TN stands for True Negative, FP stands for False Positive, and FN stands for False Negative. The proposed system attained an accuracy of 94%, indicating the overall correctness of the system's predictions. Precision is the ratio of true positive predictions (correctly predicted positive instances) to the total number of positive predictions made by the model. It gauges the accuracy of positive predictions.

$$Precision = \frac{TP}{TP + FP} \quad (7)$$

The precision of 96% indicates the proportion of correctly identified positive cases out of all predicted positives,

demonstrating the system's capacity to minimize false positives. Recall, on the other hand, is the ratio of true positive predictions to the total number of actual positive instances. It assesses the model's ability to identify all relevant instances.

$$Recall = \frac{TP}{TP+FN} \quad (8)$$

The recall rate of 87% signifies the system's capability to correctly identify most actual positive cases. F1-score, as the harmonic mean of precision and recall, offers a balanced measure between precision and recall, particularly useful when dealing with imbalanced classes.

$$F1score = \frac{2\ X\ Precision\ X\ Recall}{Precision+Recall} \quad (9)$$

The F1-score, reflecting a balance between precision and recall, was computed at 91 indicating a harmonious blend of the two metrics. Furthermore, the area under the receiver operating characteristic curve (AUC-ROC) was determined to be 93% as shown in Fig. 12.

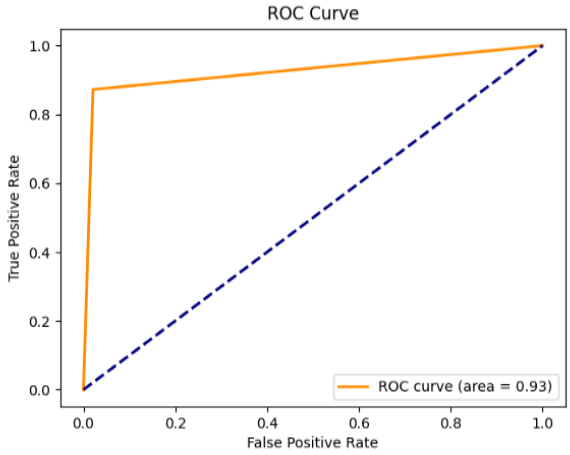

Fig. 12. ROC curve for the proposed system.

VI. Conclusion and Future Work

The project aimed to develop a comprehensive system for the detection of Polycystic Ovary Syndrome (PCOS) using a multi-modal approach, integrating both ultrasound images and clinical data. Feature extraction from ultrasound images was performed using the ResNet-50 deep learning model, achieving promising results. Three classification models-Support Vector Machine (SVM), Random Forest and Logistic Regression - were applied to classify the extracted features, with SVM demonstrating superior performance with 99% accuracy, 99% precision, 96% recall, and 98% F1 score. Furthermore, ROC curves and SHAP interpretability plots provided insights into model performance and feature importance.

In parallel, clinical data were analyzed, and feature extraction was conducted based on correlation coefficients, identifying key features relevant to PCOS diagnosis. The selected features from correlational analysis of clinical dataset were grouped into two categories of Rotterdam criteria: hyperandrogenism and Ovarian dysfunction. Decision tree was used to find if the patient has ovarian dysfunction based on the cycle length measure in days and irregularity of the cycle. Hyperandrogenism was determined by averaging different features contributing to it. Late fusion model is used to combine the results of all three deciding factors of PCOS as per Rotterdam criteria. The proposed system achieved 94% accuracy, 96% precision, 87% recall, 91% F1 score when evaluated against patients' actual health status. Thus, using two models of dataset, ultrasound images and clinical data of the patient is necessary for an accurate prediction of the syndrome. This study proves deep learning can analyze ultrasound images and clinical data for PCOS diagnosis. Enhancing the PCOS prediction through the utilization of larger, more diverse datasets containing extensive patient records encompassing various manifestations of PCOS. Incorporating additional etiological factors contributing to PCOS onset into predictive models for better forecast. Additionally, developing better methods to predict hyperandrogenism will be necessary as more features will be added to facilitate early and accurate diagnosis.

References

[1] M. Smet and A. McLennan, "Rotterdam criteria, the end," *Australas J Ultrasound Med*, vol. 21, no. 2, pp. 59–60, May 2018, doi: 10.1002/ajum.12096.

[2] M. Sumathi, P. Chitra, S. Sheela, and C. Ishwarya, "Study and implementation of automated system for detection of PCOS from ultrasound scan images using artificial intelligence," *Imaging Science Journal*, 2023, doi: 10.1080/13682199.2023.2229016.

[3] B. Purnama, U. N. Wisesti, Adiwijaya, F. Nhita, A. Gayatri, and T. Mutiah, "A Classification of Polycystic Ovary Syndrome Based on Follicle Detection of Ultrasound Images," in *2015 3rd International Conference on Information and Communication Technology (ICoICT*, 2015.

[4] Y. Deng, Y. Wang, and Y. Shen, "An automated diagnostic system of polycystic ovary syndrome based on object growing," *Artif Intell Med*, vol. 51, no. 3, pp. 199–209, Mar. 2011, doi: 10.1016/j.artmed.2010.10.002.

[5] S. S. Deshpande and A. Wakankar, "Automated Detection of Polycystic Ovarian Syndrome Using Follicle Recognition," in *2014 IEEE International Conference on Advanced Communication Control and Computing Technologies (ICACCCT)* , 2014.

[6] S. A. Suha and M. N. Islam, "An extended machine learning technique for polycystic ovary syndrome detection using ovary ultrasound image," *Sci Rep*, vol. 12, no. 1, Dec. 2022, doi: 10.1038/s41598-022-21724-0.

[7] P. Bedi, S. B. Goyal, A. S. Rajawat, and M. Kumar, "An integrated adaptive bilateral filter-based framework and attention residual U-net for detecting polycystic ovary syndrome," *Decision Analytics Journal*, vol. 10, p. 100366, Mar. 2024, doi: 10.1016/j.dajour.2023.100366.

[8] A. A. Nazarudin *et al.*, "Performance Analysis of a Novel Hybrid Segmentation Method for Polycystic Ovarian Syndrome Monitoring," *Diagnostics*, vol. 13, no. 4, Feb. 2023, doi: 10.3390/diagnostics13040750.

[9] E. Nsugbe, "An artificial intelligence-based decision support system for early diagnosis of polycystic ovaries syndrome," *Healthcare Analytics*, vol. 3, Nov. 2023, doi: 10.1016/j.health.2023.100164.

[10] S. C. R. Nandipati, X. Chew, and K. K. Wah, "Polycystic Ovarian Syndrome (PCOS) Classification and Feature Selection by Machine Learning Techniques," *Applied Mathematics and Computational Intelligence* , vol. 9, pp. 65–74, Dec. 2020.

[11] H. Elmannai *et al.*, "Polycystic Ovary Syndrome Detection Machine Learning Model Based on Optimized Feature Selection and Explainable Artificial Intelligence," *Diagnostics*, vol. 13, no. 8, Apr. 2023, doi: 10.3390/diagnostics13081506.

[12] H. Danaei Mehr and H. Polat, "Diagnosis of polycystic ovary syndrome through different machine learning and feature selection techniques," *Health Technol (Berl)*, vol. 12, no. 1, pp. 137–150, Jan. 2022, doi: 10.1007/s12553-021-00613-y.

[13] S. Bharati, P. Podder, and M. R. Hossain Mondal, "Diagnosis of Polycystic Ovary Syndrome Using Machine Learning Algorithms," in

2020 IEEE Region 10 Symposium (TENSYMP), 5-7 June 2020, Dhaka, Bangladesh , 2020.

[14] S. Nasim, M. S. Almutairi, K. Munir, A. Raza, and F. Younas, "A Novel Approach for Polycystic Ovary Syndrome Prediction Using Machine Learning in Bioinformatics," *IEEE Access*, vol. 10, pp. 97610–97624, 2022, doi: 10.1109/ACCESS.2022.3205587.

[15] D. Hdaib, N. Almajali, H. Alquran, W. A. Mustafa, W. Al-Azzawi, and A. Alkhayyat, "Detection of Polycystic Ovary Syndrome (PCOS) Using Machine Learning Algorithms," in *IICETA 2022 - 5th International Conference on Engineering Technology and its Applications*, Institute of Electrical and Electronics Engineers Inc., 2022, pp. 532–536. doi: 10.1109/IICETA54559.2022.9888677.

[16] M. Alagarsamy, N. Shanmugam, D. P. Mani, M. Thayumanavan, K. Karpoora Sundari, and K. Suriyan, "Detection of Polycystic Syndrome in Ovary Using Machine Learning Algorithm," *International Journal of Intelligent Systems and Applications in Engineering IJISAE*, vol. 2023, no. 1, pp. 246–253, 2023, [Online]. Available: www.ijisae.org

[17] A. Denny, A. Raj, A. Ashok, M. Ram C, and R. George, "i-HOPE: Detection And Prediction System For Polycystic Ovary Syndrome (PCOS) Using Machine Learning Techniques," in *2019 IEEE Region 10 Conference (TENCON 2019)*, 2019.

[18] S. Alam Suha and M. N. Islam, "Exploring the dominant features and data-driven detection of polycystic ovary syndrome through modified stacking ensemble machine learning technique," *Heliyon*, vol. 9, no. 3, Mar. 2023, doi: 10.1016/j.heliyon.2023.e14518.

[19] J. J. Cheng and S. Mahalingaiah, "Data mining polycystic ovary morphology in electronic medical record ultrasound reports," *Fertil Res Pract*, vol. 5, no. 1, Dec. 2019, doi: 10.1186/s40738-019-0067-7.

[20] V. Castro *et al.*, "Identification of subjects with polycystic ovary syndrome using electronic health records," *Reproductive Biology and Endocrinology*, vol. 13, no. 1, Oct. 2015, doi: 10.1186/s12958-015-0115-z.

[21] A. Alamoudi *et al.*, "A Deep Learning Fusion Approach to Diagnosis the Polycystic Ovary Syndrome (PCOS)," *Applied Computational Intelligence and Soft Computing*, vol. 2023, 2023, doi: 10.1155/2023/9686697.