Polycystic Ovary Syndrome Detection Using Multimodal Machine Learning

Hariharan Vijayan
MSCDAD_C_JAN25I
National College of Ireland
Dublin , Ireland
23404507

Suyash Nandkishor Shinde
MSCDAD_C_JAN25I
National College of Ireland
Dublin, Ireland
23396105

***Abstract*—Polycystic ovary syndrome (PCOS) is a hormonal disorder which affects 5-10 percent of the women between reproductive age (18-44)[4]. This paper proposes a multi modal machine learning framework that gives a diagnostic by combining Blood reports of the patients and the Ultrasound or sonographic images for robust PCOS detection. In this model we follow a basic machine learning approach for the tabular data and convolutional neural network CNN for the image data which is Ultrasound. Predicting PCOS in early stage may result in potentially reducing a long-term future risk associated with the problem. In this model we also discuss about integrating both the tabular and the image data to provide a unified predictive model.**

***Index Terms*—component, formatting, style, styling, insert**

I. Introduction

Polycystic Ovary Syndrome affects approximately 10 percent of women worldwide. It is categorized by a few symptoms namely hormonal imbalances, irregular menstruation, and cyst formation in the ovaries. The complexity of the syndrome makes it a difficult one to predict and provide early diagnostic. And the reason why the syndrome is very complex is it may not be present the same way in every patient. For example, some of them are irregularity in menstrual periods, acne, overweight, increased tendency for infertility, intense hair fall, balding of front head, increased facial hair growth [1]. Traditionally, PCOS diagnosis are mainly dependent on hormonal profiling, Clinical evaluation and ultrasounds. And performing them individually consumes time.

With the growing amounts of the electronic health records (EHRs), advanced imaging and computational tools, it is possible to use data driven approaches to augment diagnosis of PCOS. Consequently, the use of Machine learning methods, particularly, can provide accurate diagnostic support based on hidden patterns in clinical and imaging data.

In this paper, we are using a classical machine learning model to analyse blood data and deep learning to ultrasound pictures. The primary goal of this paper is to make a diagnostic approach and get a result by combining both modalities and get an end-to-end PCOS detection system.

Identify applicable funding agency here. If none, delete this.

II. Related work

The Role of Artificial intelligence in the medical and the healthcare field has been seen in an increasing trend as the medical field shifted their focus towards early intervention. Over five million women worldwide has been diagnosed with PCOS in their reproductive age . It is an endocrine disorder characterized by changes in the female hormone levels . This condition might lead to ovarian dysfunction with increased risk of miscarriage and infertility[2].hence the researchers have recently turned their attention towards machine learning and artificial intelligence to support the clinicians, With the help of these technologies, we can process large volume of clinical data and image data in a way that the diagnostic processes can't do efficiently.

The criteria for the diagnosis of PCOS were set together by the American Society for Reproductive Medicine (ASRM) and the European Society of Human Reproduction and Embryology (ESHRE). The criteria for the diagnosis of PCOS are oligo (failure to ovulate) and/or anovulation (failure to ovulate), clinical and/or biochemical signs of elevated levels of male hormones (hyperandrogenism) and polycystic ovaries on at least one of the ovaries (seen in ultrasound examination). If 2 of these 3 criteria are filled in, as recommended by the ASRM/ESHRE, PCOS should be diagnosed [3].

Currently by convention, a polycystic ovary (PCO) is described by a contour that consists of having 12 or more ovarian follicles (roughly, spherical fluid filled compartments in which eggs normally develop; their diameters range generally between 2 and 9 mm) and being responsible for more than 10 cm3 of total volume of woman ovaries.[3]

At the same time, deep learning methods have also been applied to analyze ultrasound images of ovary by parallelizing with tabular data approaches. These images usually depict a certain morphology characteristic of PCOS, for example a larger number of small follicles spread peripherally – sometimes called string of pearls pattern. However, these have been increasingly applied in this area using CNNs (Convolutional Neural Networks) - types of networks which are very good at pulling patterns out of image data.

This project also relies on these previous efforts that have been implemented the best practices in both the data streams which are structured and unstructured, but places model clarity and modularity first. It can lay the base for future more comprehensive and future ready diagnostic approach by developing special dedicated pipelines of blood report features and sonographic images.

III. DATA PROCESSING METHODOLOGY

This work presents a multimodal machine learning pipeline for the diagnosis of Polycystic Ovary Syndrome (PCOS) consisting of two leading sources of diagnostic data: clinical and biochemical data as well as ovarian ultrasound images. The variability of PCOS renders the use of one modality alone likely to result in the low accuracy and not be general. This is more in line with clinical diagnostic routines of choosing approaches that combine physiological parameters (e.g. Hormone levels and BMI) and morphological changes, as they are detected in the sonography of the ovaries. The objective is to provide an accurate and interpretable result for early intervention of PCOS.

There are three core modules.

- Tabular data-based diagnosis
- Image data-based diagnosis
- A fusion layer that combines both and produces in streamlit dashboard.

A. *Tabular data-based diagnosis*

The Data has been extracted from kaggle and it contains 3000 rows with the required columns which are needed to diagnose PCOS such as Menstrual irregularity , testosterone level. We have imported 3000 more synthetic rows to test the accuracy of improve the performance of the model. PCOS can be determined using Rotterdam criteria.

1) **Rotterdam criteria:**
   Rotterdam criteria is a diagnostic guideline used to determine Polycystic Ovary syndrome. The patient is diagnosed with PCOS if the patient has at least 2 of the below 3 features.
   a) Menstrual Cycle More than 35 days
   b) Increased Testosterone levels
   c) Polycystic Ovaries in Ultrasound

   Features:
   a) Age: The Age of the patients ranging from 18-45.
   b) BMI: The Body Mass Index, which is a measure of body fat based on height and weight, ranging from 18 to 35.
   c) Testosterone Level (ng/dL): The level of testosterone in the patient's blood, an important hormonal indicator of PCOS, ranging from 20 to 100 ng/dL.
   d) Antral Follicle Count: The number of antral follicles detected during an ultrasound, ranging from 5 to 30, which helps in assessing ovarian reserve and PCOS presence.

   Target variable is PCOS diagnosis and that is used to train the model. The Rotterdam Criteria was established in 2003 by an international consensus of experts.
2) **Data Cleaning:**
   The class imbalance was sorted using balanced sampling. The correlation matrices in conjunction with pairplots served to study feature interactions.
   Feature engineering:
   a) The features were standardized using standard scalar to normalize their distributions to zero mean and unit variance. This is crucial for SVM and logistic regression.
3) **Model Selection & Training:**
   Several classifiers were constructed:
   a) **Logistic Regression**:
   The logistic regression is one of the most important and widely used statistical method. It uses the probability of the dependent variable as the function of the independent variable. It uses sigmoid function to limit the predicted values between 0 and 1.
   With the logistic regression being simple and readable it is widely used in the clinical field. In the context of the PCOS detection, it explains data scientists and clinicians on how specific attributes like BMI and Age contribute in diagnosing PCOS.
   b) **Random Forest**:
   Random forest is a learning algorithm which uses that merges the predictions of multiple decision trees to enhance the model performance. It is operated by making numerous trees amidst of training and each uses randomly selected subsets of features and samples training. The final outcome is made with the predictions of the individual trees which is typically by majority voting. Random forest can be significant in the medical field due to its ability to add noisy data and missing data.
   c) **Support Vector Machine (SVM)**:
   The main objective of the support vector machine is to be used for classification tasks and the primary goal is to find the best possible boundary that separates the data points from different class. Through SVM medical personnel can detect PCOS between patient groups using subtle and non-linear pattern analysis of clinical features like hormone levels and follicle count.

   Hyperparameters (C, max_depth, kernel type, etc.) were adjusted using GridSearchCV with 5-fold cross-validation.

4) **Model Evaluation:**
Models were tested on an independent test set using:
   a) **Confusion Matrix**:
   It is a foundation tool and its purpose is to determine the false positive and false negative and True positive and negative values.
   b) **Precision, Recall, Accuracy, F1-score**:
   The Key indicators which are derived from the confusion matrix are Precision, Recall, accuracy and F1 score. This gives the understanding of the classification behaviour of the model. Mainly to identify PCOS and NON PCOS cases.
   c) **ROC-AUC score**:
   The Receiver Operating Characteristic - Area Under the Curve (ROC-AUC) score is calculate the model's ability to differentiate between the classes across all the threshold levels.

The best model was pickled and saved for future work in the fusion interface.

B. Image data-based diagnosis:

1) **Dataset preparation**:

Ultrasound is one of the significant clinical tools for diagnosing PCOS for its ability to reveal the volume of the Ovary , follicular distribution, and stromal density. In this study we have Greyscale Ovarian ultrasound image which are categorized into two folders with names PCOS and Non PCOS and the filenames are hidden for privacy. These images are loaded using Tensor flow's image_dataset_from_directory function which assigns labels according to the names of the folder. These datasets were divided into 3 to ensure balance

- Training(70%)
- Testing(15%)
- Validation(15%)

2) **Preprocessing** :

The images were resized to 224*224 to meet the requirements of the modern convolutional neural network (CNN) architecture. We normalized pixel values by dividing them by 255 to make them a range of 0 to 1. Data augmentation techniques like horizontal flipping, rotation and zooming are used to improve the generalization of the model and generate a more realistic variation from the imaging as optional techniques in future versions.

The data pipeline tools in TensorFlow were used to batch and prefetch images to achieve efficient usage of the GPU during training.

3) **Convolutional Neural network Architecture:** :

The convolutional neural network(CNN) is developed using Keras Sequential API. There are :

- Three convolutional blocks and the size of the block's filter increase from 32-128 which are then followed by ReLU activation and Maxpooling2D to reduce the spatial dimensions
- In order to mitigate overfitting the dropout layers of rate 0.5 are added between the dense layers
- A flatten layer to convert 2D feature maps to a 1 D vector
- A chain of 2 Dense layers with the last one having sigmoid activation for binary classification.

The primary objective of this architecture is this offer a balance between interpretability and complexity

4) **Training protocol**:
   - Adam optimizer is used in this model with the learning rate of 0.01
   - Training was conducted over 20-30 epochs. Early stopping is enabled based on the validation loss to prevent the model to be overfitted. Training /Validation Accuracy ROC- AUC were monitored Realtime

5) **Evaluation**: The test set was used for the post training evaluation which consist of:
   - Accuracy
   - Precision-Recall and F1 score
   - Confusion matrix
   - ROC-AUC Curve

C. Fusion layer: Streamlit dashboard

1) **Design**:

   - Multimodel aims to combine the two perspectives.
   - Morphological: Captured by CNN from sonographic images
   - Biochemical and symptomatic: Captured through tabular machine learning
   - In this case, we can help compensate the weakness of the one with the strength of the other

2) **Streamlit**:

   - To make the system accessible and more readable by the clinical researchers and scientists . A streamlit-Interface has been set up.
   - Here we take in an input from the users via number widgets (BMI, Follicle counts and testosterone levels.) and there is an image upload component where

it takes image formats like .jpg .png .jpeg ultrasound images.

- The runtime loading procedure includes the CNN and Random Forest models through TensorFlow and pickle while @st.cache_resource functions as a caching mechanism for improved loading speed.

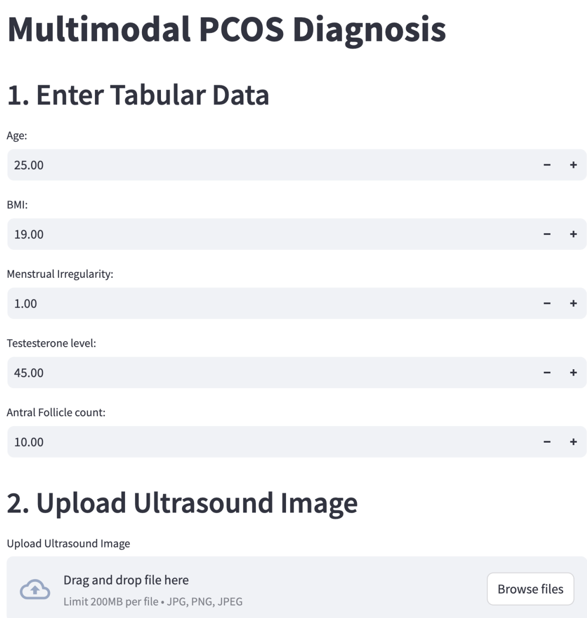

Fig. 1. Streamlit Dashboard interface

3) **Prediction strategy:**

Each model gives a output to PCOS

- Tabular data – Ptab
- Image data from the CNN model – Pimg

Now the prediction with soft fusion gives:

$$P_{\text{final}} = (P_{\text{tab}} + P_{\text{img}})/2$$

if not, it is negative

This method enables balanced contribution. And by this it wont rely entirely on a single data stream

IV. DATA VISUALIZATION METHODOLOGY

In order to explore the dataset and reveal patterns and connections among features and PCOS detection we performed EDA. They included statistical summaries(as well as visual methods such as correlation heatmaps, box plots and distribution comparisons). Patterns were to be discovered, clinically meaningful markers discovered and features selected for predictive model-building. A summary of the main findings of this analysis is presented in this Section

1) **Ultrasound**:

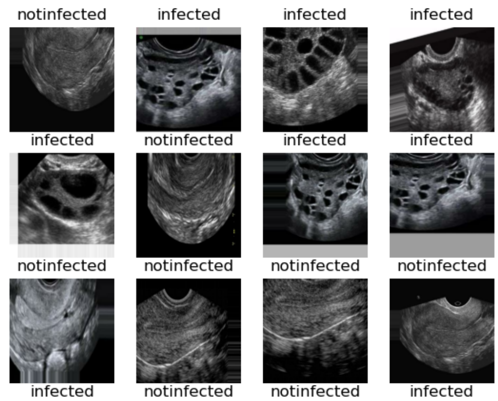

Fig. 2. Batch of image dataset

This is a figure on sample ultrasound images used for training. Observations include:

- Apparent disparities in the size, number and distribution of follicles in PCOS and non-PCOS samples.
- Stromal echogenicity and the ovarian volume which are important clinical indicators are more visible.
- Illustrates diversity in clarity and orientation of the images under strength variable, which motivates strong preprocessing and augmentation techniques.

2) **Heatmap**:

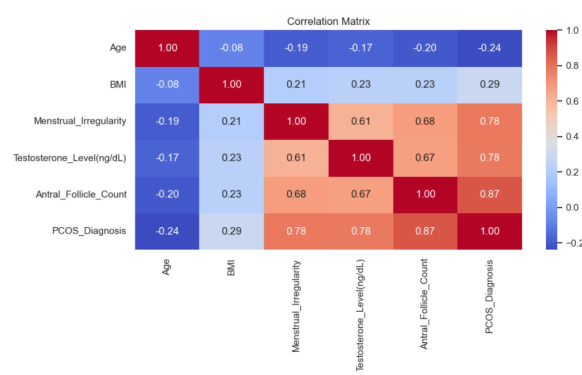

Fig. 3. Correlation analysis with heatmap

- The heatmap is a useful tool to select features: the most important thing is to balance statistic significance and use domain knowledge to have a reliable and interpretable predictive model.
- The heatmap showed a strong positive correlations between the target variable (PCOS status) and AFC, serum testosterone, and LH/FSH. In general, these findings are consistent with known clinical indicators of PCOS and suggest that these should be included as first order predictors.

3) **Antral Follicle Count (AFC) vs PCOS**

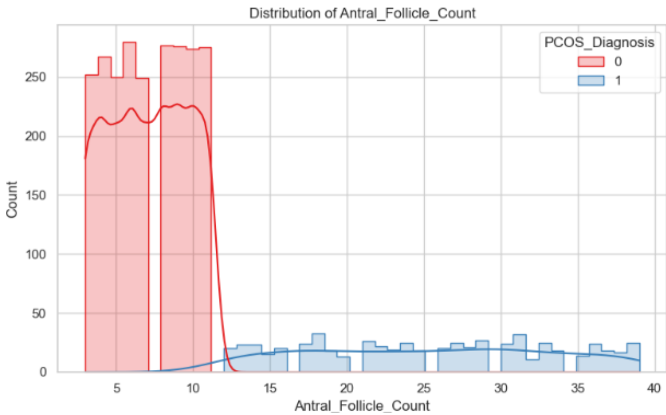

Fig. 4. Antral Follicle Count (AFC) vs PCOS

Observation:
A distinction is seen on the scatter or box plot (Fig: PCOS versus AFC) between the number of Antral Follicle Count (AFC) of women diagnosed with PCOS versus those without PCOS. Consistently, AFC values of the PCOS group are significantly above a usually applied clinical diagnostic threshold (less than 12 follicles per ovary).

Significance:
A characteristic of PCOS is hyperandrogenism, especially raised total testosterone. This causes clinical symptomatology such as hirsutism, acne, and menstrual disorders. Second, testosterone is a highly significant biomarker from a predictive modeling point of view, and it correlates very close with PCOS target label.

Significance:
A characteristic of PCOS is hyperandrogenism, especially raised total testosterone. This causes clinical symptomatology such as hirsutism, acne, and menstrual disorders. Second, testosterone is a highly significant biomarker from a predictive modeling point of view, and it correlates very close with PCOS target label.

4) **Testosterone levels vs PCOS:**

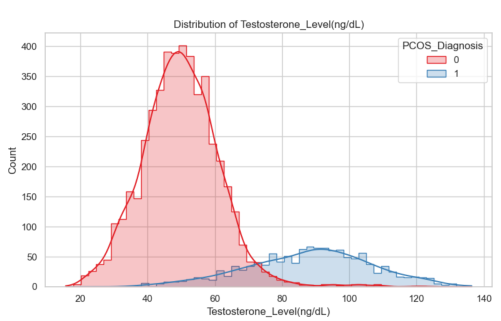

Fig. 5. Testosterone vs PCOS

Observation:
Testosterone levels are increased in PCOS patients as is shown in the plot PCOS vs Testosterone. There is indeed some overlap with non-PCOS population but the average and upper bounds are significantly larger for PCOS cases.

Significance:
A characteristic of PCOS is hyperandrogenism, especially raised total testosterone. This causes clinical symptomatology such as hirsutism, acne, and menstrual disorders. Second, testosterone is a highly significant biomarker from a predictive modeling point of view, and it correlates very close with PCOS target label

5) **BMI vs PCOS**

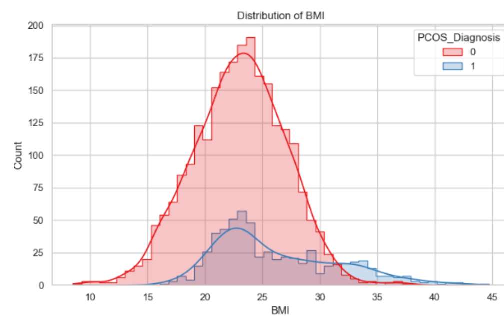

Fig. 6. BMI vs PCOS

Observation:
As shown in Figure PCOS vs BMI, Body Mass Index (BMI) values tend to be higher for individuals (of same age) who also have PCOS. Despite this, there is great overlap with non-PCOS patients suggesting that BMI alone does not have high specificity.

Significance:
Although PCOS has multiple causes, among them clinical obesity is a known risk factor which increases the severity of PCOS symptoms by making insulin resistance and hormonal imbalance worse. BMI is a moderate predictor of despite its clinical relevance: it can be used along with hormonal and ultrasound features.

6) **PCOS vs Age**

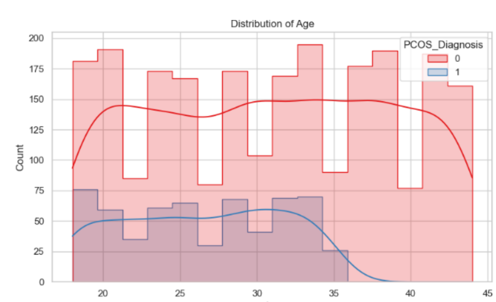

Fig. 7. PCOS vs Age

Observation:

The PCOS vs Age plot does not indicate a lot of difference in age distribution among PC0S vs non PCOS groups. The two classes do not appear to vary in mean age.

Significance:

PCOS usually occurs in reproductive years (15–44 years) but age alone does not distinguish an affected from an unaffected individual in this dataset. Furthermore, age is a weak predictor of doloknep, by itself, and should be used very carefully in training models.

V. Results

Performance Analysis

The performance of every classifier for the evaluation results is very good, however, Random Forest performs relatively very well with really high percentage of accuracy (F1-scores: 0.99 for Class 0, 0.95 for Class 1). The relatively high success of all the models means that the training data set was very clean and well structured, and thus patterns were learned well. Logistic Regression also showed a comparatively balanced precision-recall tradeoff (differences greater than or equal to 0.05), which might partly be attributable to this data quality, while SVM performed as well as it did; notwithstanding the Class 1 F1-score from SVM being much lower (0.85) than for Class 0 (0.94). The performance patterns correlate with both the dataset's inherent class imbalance (950 Class 0 vs 250 Class 1 samples) and its overall cleanliness.

Model Comparison

The best results are obtained with the random forest, which has the smallest precision-recall gaps and works well for both classes. The great results (also of Logistic Regression that performs close to this algorithm) and the good meta-classification for SVM confirms the quality of the dataset that allowed a good training. Although the results show that Random Forest works best for our problem, the fact that the performance of all classifiers (except for SVM Class 1) is very close, the 'nice' data provided a good base for all classifiers to learn meaningful patterns regarding the classification, while the architectural differences between them is responsible for the performance differences observed for other classifiers.

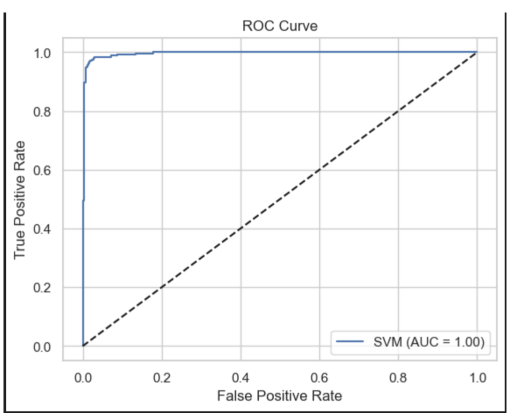

Fig. 8. ROC Curve

X-Axis:False positive rate

Y-Axis:True positive rate

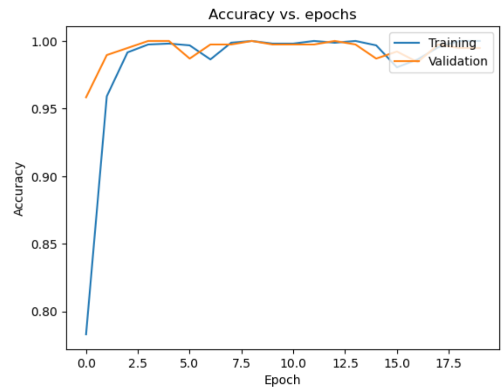

Fig. 9. Accuracy vs Epoch

This is the learning progression of the model represented on this graph. The training and validation accuracy curves climb steadily and plateau. If the two curves are close then this means that the model doesn't overfit and generalizes well. The CNN's functional efficacy on learning patterns related to PCOS from sonographic images is confirmed by this result.

- **X-Axis**:Epoch
- **Y-Axis**:Accuracy

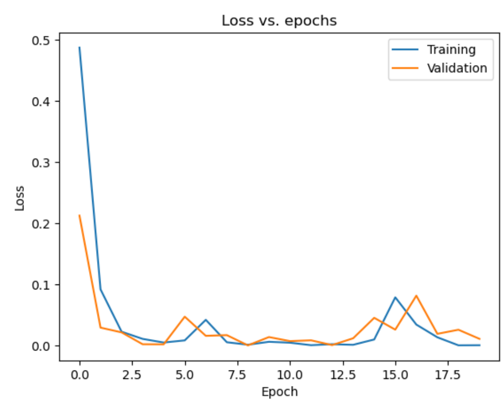

Fig. 10. Loss vs Epoch

This is how the model minimalizes its error and this is the graph presented. As around 20 epochs, the training loss and validation loss decrease steadily and stabilize. This further affirms that the model learns without memorizing training data, as the minimal gap between the two curves.

- **X Axis**: Epoch
- **Y axis**: Loss

VI. CONCLUSION

.

PCOS is the most common endocrine disorder in women of reproductive age[5] early intervention of the said disease would benefit a large amount .Systematic reviews highlight high sensitivity and accuracy of AI/ML for PCOS diagnosis and classification, emphasizing the potential for improved clinical diagnostic[6][7] The potential in combining image-based deep learning and tabular data-based machine learning in the context of Polycystic Ovarian Syndrome (PCOS) diagnosis was realized in the project by employing a multimodal pipeline in a unified web-based diagnosing web application. The convolutional neural network learned visual patterns from the ovarian ultrasonographic images, and conventional classifiers, such as Random Forests and SVM, focused on structured clinical and biochemical features. The fusion regularized—using a simple average of predicted probabilities—contributed in improving the diagnostic stability by taking advantage of both visual and numerical evidences. The end system implemented with Streamlit provides an intuitive UI for interactive PCOS risk estimation, which hopefully will help improve both clinicians and patients with PCOS.

VII. FUTURE WORK

To further enhance the diagnostic accuracy, robustness, and interpretability of the system, the following aspects can be explored:

- **Advanced Fusion Methods:**
  Replace the straightforward averaging with a trainable meta-model (e.g., logistic regression or neural network) to estimate the optimal fusion weights based on validation data.
- **Transparency AND Explainability:**
  Create explanations to increase prediction interpretability using model explanation tools such as Grad-CAM (for CNN) or SHAP (for tabular models).
  Feed back visually which part or feature of an image has most influenced the diagnosis.
- **Greater and More Varied Quantity of Data:**
  Increase the heterogeneity of ethnicity, demographics, and clinical information by including larger multicenter studies to enhance model generalizability.
  Increase the amount of data via synthetic methods or image generation techniques (GAN-based).
- **Time-Series and Longitudinal Data:** Extend the present tabular model to arrive at the inclusion of temporal data, much more like logs of the menstruation cycle or time trends of hormones and thus enhance prediction in the gray zone.
- **Clinical Application in Real Time:** Develop security options (e.g., user verification, security field).
  Validate the tool in a real-world clinical environment with feedback from gynecologists or endocrinologists..
- **Mobile and Edge Deployment:** Port the app to mobile or to edge (to run in a rural or resource limited setting) using something like TensorFlow Lite.

REFERENCES

[1] J. J. Cheng and S. Mahalingaiah, "Data mining and classification of polycystic ovaries in pelvic ultrasound reports", *bioRxiv*, pp. 254870, 2018.

[2] i-HOPE: Detection And Prediction System For Polycystic Ovary Syndrome (PCOS) Using Machine Learning Techniques.

[3] A. H. Balen, J. S. E. Laven, S.-L. Tan, and D. Dewailly, "Ultrasound assessment of the polycystic ovary: international consensus definitions", *Human Reproduction Update*, 9(6):505–514, 2003.

[4] S. Franks, "Medical progress: Polycystic ovary syndrome", *New England Journal of Medicine*, 333(13):853–861, 1995.

[5] N. F. Goodman *et al.*, "American Association of Clinical Endocrinologists American College of Endocrinology and androgen excess and PCOS society disease state clinical review: guide to the best practices in the evaluation and treatment of polycystic ovary syndrome—part 1", *Endocrine Practice*, vol. 21, no. 11, pp. 1291–1300, 2015.

[6] F. J. Barrera *et al.*, "Application of machine learning and artificial intelligence in the diagnosis and classification of polycystic ovarian syndrome: a systematic review", *Frontiers in Endocrinology*, vol. 14, p. 1106625, Sep. 2023. https://doi.org/10.3389/fendo.2023.1106625.

[7] K. Upreti, J. George, S. Upreti, and S. Mahajan, "Polycystic Ovary Syndrome Diagnosis: The Promise of Artificial Intelligence for Improved Clinical Accuracy," Institution, City, State, Country, Year.

.