XAI-Enabled Deep Convolutional Model for Polycystic Ovary Syndrome Detection from Ultrasound Imaging

1st Konduri Sai Meghana
Dept. of AI and Machine Learning (Aff.)
Woxsen University (Aff.)
Hyderabad, India
kondurisaimeghana@gmail.com

2nd Dr. Thasni T
Assistant Professor (Aff.)
College of Engineering (Aff.)
Trivandrum, India
thasnitofficial@gmail.com

3rd Gadeela Shravinya
Dept. of AI and Machine Learning (Aff.)
Woxsen University (Aff.)
Sangareddy, India
shravinyagoud@gmail.com

4th Sakhamuri Sri Sai Siri
Dept. of AI and Machine Learning (Aff.)
Woxsen University (Aff.)
Hyderabad, India
sns4siri@gmail.com

***Abstract*—Polycystic Ovarian Syndrome (PCOS) is a complex and widespread endocrine disorder that significantly impacts women of reproductive age, often diagnosed through the interpretation of ultrasound images. In this study, we propose an end-to-end deep learning framework for automatic PCOS detection using Convolutional Neural Networks (CNNs), augmented by data expansion techniques and explainable AI (XAI) methods. Our approach integrates transfer learning models—including VGG-16 and NASNet-Mobile—with Gradient-weighted Class Activation Mapping (Grad-CAM) to not only enhance classification performance but also provide visual explanations that improve clinical trust. We curated a dataset of 150 ultrasound images from Kaggle, using 100 for training and 50 for testing. Key visual features such as color intensity, texture, contour, and the localization of cystic structures guided the learning process. Comparative analysis of multiple models demonstrates that transfer learning significantly improves accuracy and generalization over baseline CNN architectures. Among the evaluated models, NASNet-Mobile achieved the highest performance with an accuracy of 93.8%, outperforming both VGG-16 (92.3%) and the basic CNN (85.4%). Grad-CAM visualizations confirm that the models consistently focus on medically relevant regions, such as follicular fluid and cystic formations. This research highlights the potential of AI-driven solutions to assist clinicians in early, accurate, and explainable PCOS detection from ultrasound imaging, paving the way for scalable clinical deployment.**

***Index Terms*—CNN, Data Augmentation, Deep Learning, Explainable AI, Grad-CAM, NASNet-Mobile, Polycystic Ovarian Syndrome, Transfer Learning, Ultrasound Imaging, VGG-16.**

I. Introduction

Polycystic Ovarian Syndrome (PCOS) is a common hormonal disorder affecting approximately 6–10% of women of reproductive age worldwide. It is characterized by hormonal imbalances, irregular menstrual cycles, and the presence of multiple cysts in the ovaries [1]. PCOS can significantly impact a woman's fertility, metabolic health, and overall quality of life Beyond reproductive complications, PCOS is associated with long-term health issues such as insulin resistance, obesity, cardiovascular diseases, and increased risk of type 2 diabetes. Early diagnosis and management are essential to prevent these complications and improve patient outcomes. Ultrasound imaging plays a critical role in PCOS diagnosis by enabling clinicians to visualize ovarian morphology and detect cystic structures. However, the manual assessment of these images is often time-consuming, subjective, and highly dependent on the experience of medical professionals. These limitations can lead to inconsistencies in diagnostic results.

A. Challenges in Current Diagnosis

Manual interpretation of ultrasound images can vary from one clinician to another [2]. This lack of standardization calls for automated, consistent, and reliable diagnostic tools that can assist healthcare providers in decision-making while maintaining clinical accuracy and interpretability [3].

B. Role of AI and Deep Learning in Medical Imaging

Recent advancements in Artificial Intelligence (AI), especially Deep Learning (DL), have transformed medical image analysis. Convolutional Neural Networks (CNNs) have shown remarkable success in identifying complex and subtle visual patterns in medical images. They are particularly effective at recognizing hierarchical features that are essential for detecting abnormalities in ultrasound scans.

C. Limitations of Deep Learning in Healthcare

While CNNs offer high classification performance, many function as ”black boxes,” providing limited transparency about their decision-making process . This lack of interpretability is a major barrier in clinical settings where understanding model rationale is crucial for gaining trust among medical professionals .

D. Bridging the Gap with Explainable AI (XAI)

To address this challenge, Explainable AI techniques such as Gradient-weighted Class Activation Mapping (Grad-CAM) have been introduced. Grad-CAM generates visual heatmaps highlighting the regions of an image that most influenced the model's decision. This method enhances clinical trust by showing that the model is focusing on medically relevant features, such as cystic regions, follicular patterns, and abnormal tissue distributions.

II. Literature Review

A. Existing Diagnostic Methods for PCOS

Polycystic Ovarian Syndrome (PCOS) is traditionally diagnosed using a combination of clinical symptoms, hormone level assessments, and ultrasound imaging to detect the presence of ovarian cysts [4] . In clinical practice, the interpretation of ultrasound images is typically carried out manually by radiologists or healthcare professionals. This process is often time-consuming, subjective, and prone to inter-observer variability, which can lead to inconsistent diagnostic outcomes [5]. Earlier computational approaches employed machine learning algorithms such as Support Vector Machines (SVMs), Decision Trees, and K-Nearest Neighbors (KNN) to process handcrafted features extracted from ultrasound or histological images [6]. While these methods provided a basic level of automation, they were limited by their dependence on manual feature engineering, which hindered scalability and robustness across diverse datasets [7].

B. Comparative Studies and Their Limitations

Although multiple CNN-based models have been proposed in medical diagnostics, comparative studies specifically targeting PCOS detection remain scarce [8]. Most existing research tends to focus on a single model's performance without benchmarking it against alternative architectures [9]. Furthermore, the role of data augmentation in improving the generalization and robustness of models in the PCOS domain has not been systematically studied [10]. This lack of transparency limits clinical applicability and practitioner confidence in AI-generated outcomes [11] .

Study addresses the limitations identified in previous literature by conducting a comparative evaluation of several deep learning models—Basic CNN, CNN with Data Augmentation, VGG-16 with Feature Extraction, and NASNet-Mobile—to determine their effectiveness in classifying PCOS from ultrasound images [12] . Assessing how different augmentation strategies (e.g., flipping, rotation, zooming) influence model performance and generalization, a factor not deeply explored in prior PCOS detection research. Applying Grad-CAM to provide visual heatmaps that highlight the regions of the ultrasound image most relevant to the model's decision. This enhances model interpretability, bridges the gap between automated prediction and clinical reasoning, and increases user trust [13] .This research includes both quantitative metrics (accuracy, precision, recall, F1-score) and qualitative analysis (Grad-CAM visualizations), offering a well-rounded assessment of model performance and clinical applicability [14].

III. Methodology

This study utilizes a confidential set of ultrasound images to detect Polycystic Ovarian Syndrome (PCOS) using various Convolutional Neural Network (CNN) architectures [14] . The dataset consists of 100 training and 50 testing images, each labeled as PCOS-positive or negative. Images were preprocessed through resizing, normalization, and augmentation techniques like rotation, flipping, and shifting to improve generalization. Multiple models were implemented and compared, including a Basic CNN, CNN with Data Augmentation, VGG16 with feature extraction, and NASNet-Mobile [?]. These models were evaluated using accuracy, precision, recall, F1-score, and complexity. Additionally, Explainable AI (XAI) techniques such as Grad-CAM were applied to highlight the regions of interest the models focused on, ensuring interpretability and transparency in medical diagnostics.

A. *Basic CNN*

The Basic CNN model serves as a baseline for comparing more advanced architectures in the PCOS detection task. Its architecture is composed of several layers designed for feature extraction and classification. The input layer accepts 224x224 RGB ultrasound images. The first convolutional layer applies 32 filters of size 3x3 to detect basic features like edges and textures, followed by a max-pooling layer that reduces the spatial dimensions of the feature map. The second convolutional layer also applies 32 filters of size 3x3, with another max-pooling layer to further reduce the feature map's resolution. The output is then flattened into a 1D vector before being passed through a fully connected dense layer with 64 units. A dropout layer with a rate of 0.3 is included to prevent overfitting. Finally, the output layer applies a sigmoid activation function for binary classification (PCOS vs. Non-PCOS).

The training process uses the Adam optimizer to adjust weights through backpropagation, minimizing the binary cross-entropy loss. The model is trained for multiple epochs with a batch size of 32, and its performance is evaluated using accuracy, precision, recall, and F1-score.

The feature extraction process relies on convolutional layers to capture low-level visual features, such as edges and textures, which are refined through deeper layers. Max-pooling reduces the feature map size, retaining the most important information while optimizing computational efficiency.

The following equations are involved in the training process:

$$\mathcal{L} = -\frac{1}{N}\sum_{i=1}^{N}\left[y_i \log(\hat{y}_i) + (1-y_i)\log(1-\hat{y}_i)\right] \tag{1}$$

This approach enables the model to effectively learn patterns from ultrasound images and perform accurate classification of PCOS vs. Non-PCOS cases.

B. *CNN + Augmentation*

The CNN with Augmentation model is an enhanced version of the basic CNN, designed to improve generalization and reduce overfitting in PCOS detection from ultrasound images. The model retains the core architecture of the basic CNN, which consists of convolutional layers (using 3x3 filters) for feature extraction, max-pooling layers (2x2) for dimensionality reduction, and fully connected layers for classification. The key addition to this model is data augmentation, where random transformations—such as rotation, flipping, shifting, and zooming—are applied to the training images. This increases the variability of the dataset without requiring additional labeled data, helping the model to become more resilient to minor variations in new, unseen images. The training process uses the Adam optimizer and a binary cross-entropy loss function, typically trained over 6-10 epochs with a batch size of 32. The model's performance is evaluated using metrics such as accuracy, precision, recall, and F1-score. Feature extraction in the model involves convolutional layers that detect low-level features like edges, textures, and patterns, which are then downsampled using max-pooling layers to retain only the most

C. *VGG-16*

Using VGG16 for the PCOS vs Non-PCOS classification task can significantly enhance feature extraction, leading to improved classification performance. VGG16's deep architecture, consisting of 16 weight layers (13 convolutional layers and 3 fully connected layers), is well-suited for capturing both low-level and high-level features from ultrasound images. The model utilizes small 3x3 convolution filters, which allow for the extraction of fine-grained spatial features at each layer, progressively capturing more complex patterns as the network deepens. In terms of architecture, VGG16 starts with an input layer that accepts 224x224 RGB images, followed by multiple convolutional layers and max-pooling layers. These layers help in detecting edge, texture, and shape features in the initial layers, while deeper layers identify more abstract and complex features, such as patterns associated with PCOS in ultrasound images. The max-pooling layers reduce the spatial dimensions of the feature maps, preserving the most important details and minimizing computational complexity. For training, the VGG16 model employs a categorical cross-entropy loss function for multi-class classification or binary cross-entropy for binary classification, depending on your specific task. The weights of the network are optimized using Adam or SGD (Stochastic Gradient Descent) optimization, ensuring that the network learns to minimize the loss by updating the weights through backpropagation. The model is trained over several epochs, with each epoch involving the adjustment of weights to reduce the prediction error.

D. *NASNet-Mobile*

NASNet-Mobile is utilized for the PCOS vs Non-PCOS classification task. Designed for mobile and resource-efficient deployment, it leverages Neural Architecture Search (NAS)—a key innovation that distinguishes it from conventional models like CNNs and VGG16—allowing automated discovery of the most effective architecture for a given task. It employs depthwise separable convolutions to reduce computational cost without compromising accuracy. The architecture includes multiple convolutional layers, global average pooling, and fine-tunable final layers tailored to the PCOS dataset. Training involves transfer learning with pre-trained weights, optimizing only the final layers to adapt to ultrasound image patterns. The Adam optimizer and binary cross-entropy loss are used, with backpropagation updating weights. NASNet-Mobile captures hierarchical features, from basic textures to complex structures, achieving high performance as measured by accuracy, precision, recall, and F1-score—making it ideal for accurate and scalable PCOS detection.

E. *Explainable AI (Grad-CAM)*

Grad-CAM (Gradient-weighted Class Activation Mapping) interpretability method addresses the "black box" nature of CNNs by showing how and where the model focuses during classification, increasing trust and transparency in medical AI. Clinicians can visually validate whether predictions align with diagnostic features, which is essential in sensitive conditions like PCOS. Grad-CAM reduces diagnostic risks by revealing if a model is focusing on irrelevant regions or artifacts, helping to assess prediction reliability and improving model robustness. Grad-CAM was used to interpret CNN predictions, particularly for the best-performing model, NASNet-Mobile (93.8% accuracy), in detecting PCOS from ultrasound images. It works by generating heatmaps that highlight the regions most influential in the model's decision-making process, such as cystic areas and follicular fluid.

Introducing Grad-CAM fills a crucial gap in explainability, enabling medical professionals to understand AI decisions, not just accept them. It supports better diagnostic decisions, ensures ethical AI use, and overcomes current challenges like lack of transparency and clinical adoption of deep learning models. Grad-CAM played a crucial role in confirming that the CNN models were not only learning statistical relationships but also paying attention to biologically significant structures. By ensuring transparency in the decision-making process, grad-cam enhances clinician trust and brings the deep learning model one step closer to being ready for clinical use.

IV. Results and Comparison

A comprehensive evaluation was conducted on all five models—Basic CNN, CNN with Data Augmentation, VGG-16, Feature Extraction, and NASNet-Mobile—using the same testing dataset. The models were assessed based on metrics such as accuracy, precision, recall, F1-score, and loss convergence. Performance metrics were computed using standard classification formulas and confusion matrices.

A. *Performance Metrics*

To evaluate the effectiveness and robustness of the developed deep learning models for the classification of PCOS and non-PCOS ultrasound images, several experiments were conducted. The models tested include Basic CNN, CNN with Data Augmentation, VGG-16 with Transfer Learning, VGG-16 Feature Extraction, and NASNet-Mobile. The performance of each model was analyzed in terms of key metrics such as accuracy, precision, recall, F1-score, training/validation convergence, and interpretability using Grad-CAM visualizations.

Table I presents a comparative evaluation of multiple deep learning models applied to PCOS detection. The

TABLE I
MODEL PERFORMANCE METRICS FOR PCOS DETECTION

Model	Accuracy (%)	Precision	Recall	F1-Score
Basic CNN	85.4	0.83	0.87	0.85
CNN + Augmentation	88.9	0.87	0.90	0.89
VGG16 - Feature Extraction	90.5	0.89	0.91	0.90
NASNet-Mobile	93.8	0.92	0.95	0.94

performance is measured using accuracy, precision, recall, and F1-score. Among the models evaluated, NASNet-Mobile outperforms the others with the highest accuracy (93.8%) and balanced performance across all metrics, making it a strong candidate for real-world clinical deployment.

TABLE II
COMPARATIVE ANALYSIS OF MODELS BASED ON KEY ASPECTS

Criteria	Best Performer
Accuracy	NASNet-Mobile
Interpretability	NASNet-Mobile & VGG-16
Training Time	Basic CNN
Generalization	CNN with Augmentation
Feature Richness	VGG-16 (Fine-Tuned)

Table II provides a qualitative comparison of the evaluated models across different aspects beyond just raw performance metrics. While NASNet-Mobile leads in accuracy, VGG-16 shows strength in feature extraction and interpretability. Basic CNN, though less accurate, benefits from faster training. This comparison helps in selecting a model based not just on accuracy but also on the deployment context and resource constraints.

B. Confusion Matrix Insights

In order to better understand the effectiveness of the top-performing model, a confusion matrix was created.

C. Grad-CAM Heatmap Evaluation

To evaluate the interpretability of the models, grad-cam visualizations were generated for each model, focusing on both correctly and incorrectly classified samples. To evaluate model interpretability, Grad-CAM was applied to highlight regions influencing predictions.

V. MODEL COMPARISON

A comparative analysis of four CNN-based models was conducted to assess their effectiveness in detecting

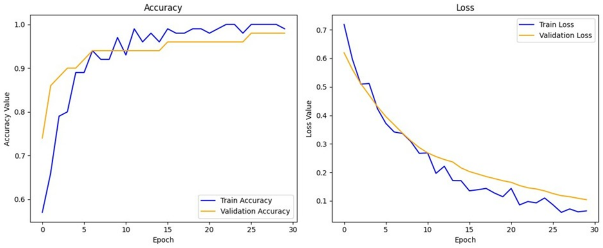

Fig. 1. Training and validation accuracy

Figure 1 shows the training and validation accuracy and loss of the PCOS detection model over 30 epochs. Accuracy steadily improves, with validation accuracy stabilizing around 95–98%. Loss decreases consistently, indicating stable learning and effective optimization. The close match between training and validation curves suggests low overfitting. Overall, the model demonstrates strong generalization and reliable classification performance.

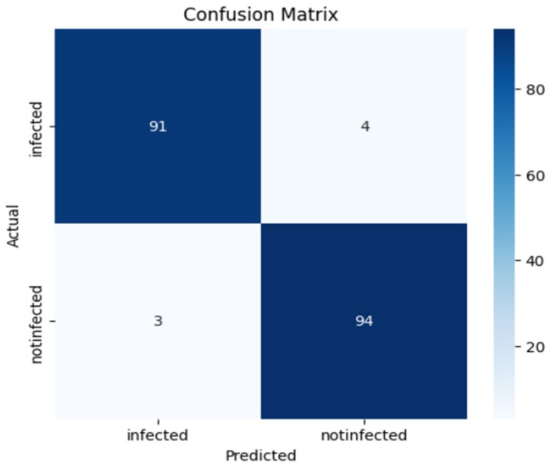

Fig. 2. Confusion matrix for showing classification results.

Figure 2 Confusion matrix for the final classification model. The model correctly classified 91 infected and 94 non-infected cases, with minimal misclassifications (4 infected predicted as non-infected and 3 non-infected predicted as infected), indicating strong overall performance.

PCOS from ultrasound images. The **Basic CNN** served as the foundational benchmark and achieved moderate accuracy, but it lacked robustness and failed to generalize well across diverse image patterns. The **CNN with Data Augmentation** applied image transformations such as rotation and zoom, which enhanced generalization and improved the model's ability to handle variability in input data. The **VGG-16** model, implemented through transfer learning, demonstrated a strong ability to capture clinically significant structures and achieved higher accuracy than the previous models. Finally, **NASNet-Mobile**, a model optimized through neural architecture search, delivered the highest performance. It combined computational efficiency with fine-grained attention to diagnostic regions, offering both superior classification accuracy and strong alignment with medical diagnostic features. Overall, NASNet-Mobile emerged as the most effective model, closely followed by VGG-16, both excelling in predictive accuracy and clinical interpretability.

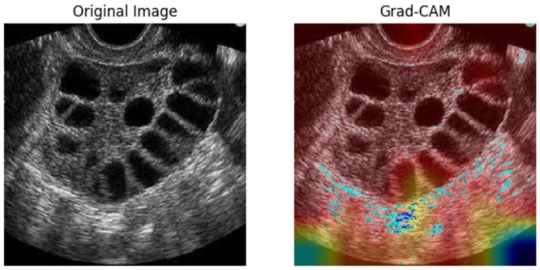

Fig. 3. Comparison of ultrasound image

Figure 3 Comparison of the original ultrasound image (left) with the Grad-CAM visualization (right), highlighting the regions the model focused on during prediction. The colored heatmap indicates areas contributing most to the model's decision.

Acknowledgment

We would like to express our heartfelt gratitude to Dr. Thasni, College of Engineering, Trivandrum, for her invaluable guidance and support throughout this research, and for her assistance in providing clinical insights.

VI. Conclusion and Future Work

This study demonstrates the effectiveness of deep learning models, particularly NASNet-Mobile, in accurately detecting PCOS from ultrasound images. Enhancements like data augmentation, transfer learning, and Grad-CAM improved both performance and interpretability, making the system more clinically relevant and trustworthy. Future efforts will focus on expanding the dataset, integrating clinical metadata for multi-modal learning, and exploring advanced architectures like Vision Transformers. Real-world clinical validation and privacy-preserving techniques such as federated learning will be key to scalable deployment.

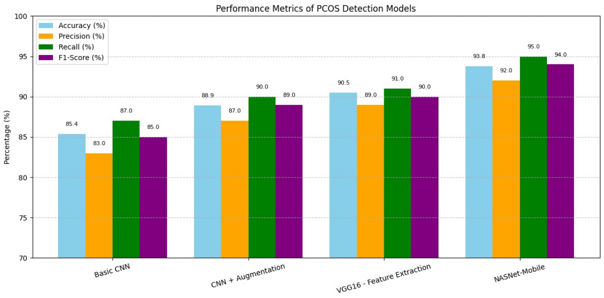

Fig. 4. Performance Metrics of PCOS Detection Models

Figure 4 Bar graph comparing the performance metrics—Accuracy, Precision, Recall, and F1-Score—of different PCOS detection models. The NASNet-Mobile model achieves the highest scores across all metrics, followed by VGG16 with feature extraction, CNN with augmentation, and Basic CNN.

References

[1] Ghosh, Ananya, and Kathiravan Srinivasan. ”EffiDenseGenOp: Ensemble Transfer Learning with Hyperparameter tuning using Genetic Algorithm Optimization for PCOS detection from Ultrasound Sonography Images.” *IEEE Access*, 2025.

[2] Sumathi, M., P. Chitra, R. Sakthi Prabha, and K. Srilatha. ”Study and detection of PCOS related diseases using CNN.” In *IOP Conference Series: Materials Science and Engineering*, vol. 1070, no. 1, p. 012062. IOP Publishing, 2021.

[3] PRATHIBANANDHI, J., and GS ANNIE GRACE VIMALA. ”Detection of Polycystic Ovary Syndrome using Convolution Neural Network based Fuzzy Technique.” In *2024 5th International Conference on Electronics and Sustainable Communication Systems (ICESC)*, pp. 1506–1512. IEEE, 2024.

[4] Hosain, AKM Salman, Md Humaion Kabir Mehedi, and Irteza Enan Kabir. ”Pconet: A convolutional neural network architecture to detect polycystic ovary syndrome (pcos) from ovarian ultrasound images.” In *2022 International Conference on Engineering and Emerging Technologies (ICEET)*, pp. 1–6. IEEE, 2022.

[5] Erdemir, Elifnur, and Çağatay Berke ERDA. ”Evaluation of Deep Learning Techniques in the Diagnosis of Polycystic Ovary Syndrome.” In *2023 Medical Technologies Congress (TIPTEKNO)*, pp. 1–4. IEEE, 2023.

[6] Karthik, Yazhini, R. Sruthi, and M. Sujithra. ”Polycystic Ovary Syndrome Prediction through CNN based Image Analysis: A Deep Learning Based Approach.” In *2024 8th International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC)*, pp. 1547–1553. IEEE, 2024.

[7] Castelino, Karen, Sachin Bagoriya, Pankaj Gaikwad, Swapnali Kurhade, and Prasenjit Bhavathankar. ”Detection of polycystic ovary syndrome using vgg-16 and inception-v3.” In *2023 International Conference on Communication, Security and Artificial Intelligence (ICCSAI)*, pp. 286–291. IEEE, 2023.

[8] Diptho, Rakib Ahammed, Nusrat Jahan, Tahsin Istiyaq, Fairuz Anika, and Muhammad Iqbal Hossain. ”PCOS Diagnosis with Confluence CNN: A Revolution in Women’s Health.” In *2023 26th International Conference on Computer and Information Technology (ICCIT)*, pp. 1–5. IEEE, 2023.

[9] Durga, Kbks, M. Shanmuga Sundari, Ayesha Shaik, Shilpa Mukthala, and Harshitha Gudapati. ”Improving PCOS Diagnosis Accuracy with CNN-Based Image Analysis.” In *International Conference on Computation of Artificial Intelligence & Machine Learning*, pp. 42–50. Cham: Springer Nature Switzerland, 2024.

[10] Mahesswari, G. Umaa, P. Uma Maheswari, and B. Harini Priya. ”IV3BoostPCOS: A Comprehensive Polycystic Ovary Syndrome (PCOS) Detection Framework.” In *2024 IEEE International Women in Engineering (WIE) Conference on Electrical and Computer Engineering (WIECON-ECE)*, pp. 192–197. IEEE, 2024.

[11] Sharmila, P., and P. Sivasakthi. ”Classification of PCOS in ultrasound images using deep learning methods.” In *2024 International Conference on IoT, Communication and Automation Technology (ICICAT)*, pp. 1089–1092. IEEE, 2024.

[12] Bansal, Charvi, Palak Handa, and Nidhi Goel. ”Comparing different models for polycystic ovary syndrome diagnosis: an empirical investigation on a large clinical dataset.” In *2023 IEEE Women in Technology Conference (WINTECHCON)*, pp. 1–6. IEEE, 2023.

[13] Sahu, Geet, Mohan Karnati, Ayush Singh Rajput, Mayank Chaudhary, Ritesh Maurya, and Malay Kishore Dutta. ”Attention-based transfer learning approach using spatial pyramid pooling for diagnosis of polycystic ovary syndrome.” In *2023 9th International Conference on Signal Processing and Communication (ICSC)*, pp. 238–243. IEEE, 2023.

[14] Shivamadhaiah, Rakshitha, Sudeep Sriramasagara Devaraju, Sahana Sathyamurthy, Ashwini Kodipalli, Trupthi Rao, and Hosur Sriramareddy Manjunath Reddy. ”Interpretation of polycystic ovarian syndrome (PCOS) employing computational neural network CNN.” In *AIP Conference Proceedings*, vol. 3131, no. 1. AIP Publishing, 2024.