Spectrum of Engineering Sciences
ISSN (e) 3007-3138 (p) 3007-312X Volume 3, Issue 7, 2025

# A DEEP LEARNING APPROACH TO PCOS DIAGNOSIS: TWO-STREAM CNN WITH TRANSFORMER ATTENTION MECHANISM

**Tariq Rahim[1], Ibadullah[*2], Aimal Nazir[3], Muhammad Salih Tanveer[4], Muhammad Rohan Qureshi[5]**

*[1]Department of Creative Technologies, Air University, Islamabad, Pakistan.*
*[*2]Department of Computing, Riphah International University, Islamabad, Pakistan.*
*[3]University of Engineering and Technology, Mardan, KPK, Pakistan.*
*[4,5]COMSATS University, Islamabad, Pakistan.*

[*1]tariqrahim885@gmail.com , [2]ibad.ullah@riphah.edu.pk ,
[3]imalnazir@gmail.com,[4]salih.hashmi.1999@gmail.com, [5]mrquetain@gmail.com

**DOI:** https://doi.org/10.5281/zenodo.15790016

**Keywords**
*PCOS classification, Two-Stream CNN, Transformer Attention, Ultrasound imaging, Deep Learning, Multi-Head Attention, Medical image analysis*

**Article History**
*Received on 27 May 2025*
*Accepted on 27 June 2025*
*Published on 02 July 2025*


**Corresponding Author:** *
**Ibadullah**

**Abstract**
*Polycystic Ovary Syndrome (PCOS) is a prevalent endocrine disorder affecting a significant portion of women worldwide, often underdiagnosed due to the complexity of its symptoms and limitations in existing diagnostic tools. With advancements in deep learning and medical imaging, automated classification systems offer the potential to revolutionize PCOS detection through precision and scalability. This study proposes a novel Two-Stream Convolutional Neural Network (CNN) architecture enhanced with Transformer-based attention mechanisms for classifying PCOS from ultrasound images. Leveraging dataset of 11,784 images, our framework splits each ultrasound image into upper and lower halves to capture anatomical variance and apply convolutional encoding separately. A Multi-Head Attention layer then integrates spatial dependencies between the two streams, enhancing feature discrimination and improving model interpretability. Experimental evaluations show that the proposed model achieves a classification accuracy of 98.96%, an F1-score of 0.99, and minimal loss on the test dataset. These results highlight the model's robustness and potential applicability in real-world clinical settings for the early detection of PCOS.*

## INTRODUCTION

Polycystic Ovary Syndrome (PCOS) is a complex endocrine disorder affecting females of reproductive age. Its global prevalence varies between 6% and 21%. depending on population characteristics and the criteria used for diagnosis [1]. Clinically, PCOS encompasses a heterogeneous spectrum of symptoms including oligo or anovulation, hyperandrogenism, and polycystic ovarian morphology. In addition to infertility, it has been associated with metabolic disorders such as insulin resistance and type 2 diabetes mellitus, obesity, cardiovascular diseases, and an increased risk of psychological disorders including anxiety and depression [2]. Despite being clinically important, PCOS tends to go unrecognized because of the staggering lack of uniform diagnostic frameworks as well as the diverse symptomatic representation across different patients.

Ultrasound imaging remains a staple in PCOS diagnosis and treatment, particularly in the assessment of ovarian morphology. It permits clinicians to evaluate such features as the number of follicles, their distribution, and overall ovarian volume. However, this process of diagnosis is typically time-consuming and subject to considerable variability due to its human interpretation dependence. Radiologists' and gynecologists' variable interpretations, particularly in unspecialized healthcare settings with limited access to specialist healthcare experts, lead to misdiagnosis or delayed diagnosis [3]. Such inadequacies point to the urgent need for highly performing and automated image-based diagnosis systems capable of delivering high accuracy as well as consistency without regard to the clinical context.

Recent advancements in artificial intelligence, particularly in deep learning, have demonstrated promising outcomes in medical image analysis. Convolutional Neural Networks (CNNs) have emerged as a groundbreaking tool, capable of autonomously learning hierarchical features from raw imaging data, thereby eliminating the need for manual feature engineering.For disease classification, segmentation, and detection applications, CNNs have surpassed most conventional machine learning approaches [5]. Yet, traditional CNN models tend to lack the ability to capture long-range dependencies and global contextual information capacity [6] that can be used to differentiate intricate anatomical patterns, including the nuanced morphological variations seen in the ovarian structures of PCOS versus non-PCOS cases.

To address these limitations, we propose a novel deep learning framework that integrates a Two-Stream Convolutional Neural Network with Transformer-based attention mechanisms. Unlike conventional CNNs that process images as unified inputs, our architecture divides each ultrasound image into upper and lower regions to capture region-specific morphological features that might be indicative of PCOS. These regions are processed through parallel CNN pipelines that allow the model to independently extract localized features from each half. To effectively integrate spatial context and learn inter-regional dependencies, we incorporate a Transformer-based Multi-Head Attention mechanism. This attention layer facilitates dynamic weighting of features across spatial dimensions, thereby enriching the feature representation and enabling the model to better distinguish between PCOS and non-PCOS images.

This high-level design allows the network to focus both on local and global characteristics of ovarian morphology, potentially improving classification accuracy and model interpretability. Rather than relying solely on pixel-level pattern recognition, the Transformer-enhanced architecture captures abstract relationships between distant parts of the image, a capacity that is critical when morphological abnormalities are subtle or non-contiguous.

The data set used in this study comprises 13,568 labeled ultrasound images, equally divided between PCOS-positive and PCOS-negative cases. These images were then subjected to a rigorous preprocessing pipeline consisting of null and duplicate row removal, normalization, and up sampling for handling class imbalance. The dataset was then split into training, validation, and test sets through stratified sampling to make class representation equal. This dataset not only provides a robust foundation for the training of a deep learning model but also allows strict validation to evaluate model generalizability.

Rather than delving into architectural specifics in this introductory section, we emphasize the conceptual innovation and clinical significance of the approach. Detailed descriptions of the model's architecture, training parameters, and implementation specifics are provided in the Methodology section of the paper.

**The primary contributions of this study are as follows:**

1. A novel Two-Stream CNN framework tailored to analyze region-specific anatomical features in PCOS ultrasound images.
2. The integration of Transformer-based Multi-Head Attention mechanisms to capture inter-regional dependencies, enhancing the model's discriminative capabilities.
3. A comprehensive evaluation pipeline that leverages balanced datasets, data augmentation, and rigorous validation to ensure clinical applicability and model generalizability.
4. Demonstration of state-of-the-art classification performance in PCOS detection, with potential scalability to other non-contiguous disease classifications in medical imaging.

The Proposed study is structured as follows: Section 2 provides a detailed literature review on PCOS diagnosis and the application of deep learning to medical image processing. Section 3 defines the proposed methodology, including the design of the Two-Stream CNN with Transformer-based attention mechanisms. Section 4 provides the results yielded upon model training and evaluation. Section 5 summarizes the research with main findings and provides directions for future research in clinical integration and broader disease classification applications.

**2: Literature Review**

Polycystic Ovary Syndrome (PCOS) is a multifactorial endocrine disorder that presents diagnostic challenges due to its varied symptoms and overlapping clinical markers [7]. In recent years, machine learning (ML) and deep learning (DL) techniques have emerged as effective tools for identifying patterns in medical data, offering promising alternatives for early and accurate PCOS detection [8].

A web-based diagnostic tool was proposed that used multiple machine learning models, namely Logistic Regression, Decision Tree, AdaBoost, Random Forest, and Support Vector Machine [9]. Mutual Information for feature selection and robust data preprocessing were used in their methodology with a patient dataset of 541 samples. Random Forest and AdaBoost returned the best accuracy of 94% from the models tested. Notably, these models were incorporated into a web interface based on Django for real-time diagnosis usability. Likewise, [10] designed the Smart PCOS Diagnostic System (SPOSDS), with a focus on non-invasive features and model efficiency. Employing correlation-based feature selection and result validation with Out-of-Bag (OOB) error estimation, their Random Forest model achieved 93.25% accuracy, attesting to its suitability for clinical screening applications.

On the other hand [11] adopted a more holistic approach by integrating classic ML with deep learning models, such as CNN, RNN, LSTM, and BLSTM. Ensemble methods like stacking and boosting were also used in their study, resulting in a better accuracy of 99.32% with a hybrid model consisting of a boosted Random Forest combined with Support Vector Classifier. This paper emphasized the importance of sophisticated model structure and hyperparameter optimization in identifying the intricate patterns of PCOS. Overall, these three studies illustrate that AI-powered models, especially ensemble and deep learning-based models, are extremely useful for refining the accuracy, accessibility, and scalability of PCOS diagnosis.

Recent advancements in machine learning (ML) and explainable AI (XAI) have greatly improved PCOS detection. [12] proposed a robust framework combining optimized feature selection with ensemble ML models. Using a Kaggle PCOS dataset, they applied models such as logistic regression, SVM, decision tree, random forest, and XGBoost. Feature selection methods like RFE, mutual information, and tree-based filtering enhanced model performance and interpretability. They addressed class imbalance with SMOTE-ENN and tuned hyperparameters using Bayesian optimization. Their stacking ML model with RFE achieved 100% accuracy, emphasizing the importance of XAI for transparent, trustworthy predictions. This integration of performance and explainability sets a strong benchmark in AI-based PCOS diagnostics.

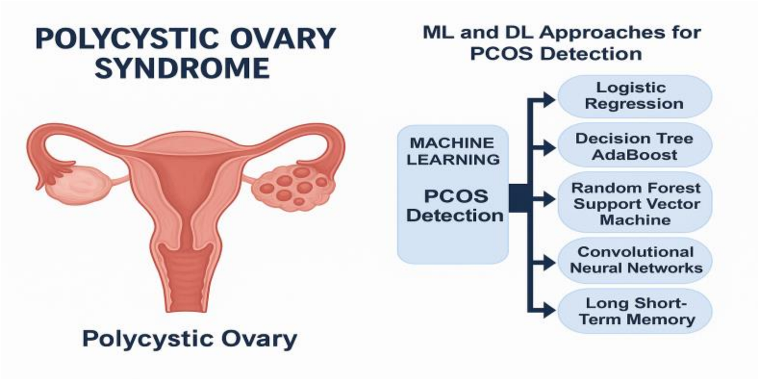


**Figure 1 PCOS overview and AI-based detection approaches.**

Figure 1 presents a visual summary of how AI techniques like ML and DL are applied to diagnose PCOS efficiently and non-invasively. [13] introduced a hybrid ML framework using Particle Swarm Optimization (PSO) to enhance PCOS diagnosis. Drawing on clinical data from ten hospitals in Kerala, they applied and optimized nine ML classifiers, including LR, SVM, KNN, RF, and XGBoost. Their LR+PSO model achieved the highest accuracy at 96.30%, with strong sensitivity (94.44%) and specificity (97.22%). The study involved detailed preprocessing, including one-hot encoding and normalization, and used correlation heatmaps for feature relevance. By comparing with previous studies, the authors demonstrated the advantages of PSO-based hybrid models for accurate, non-invasive PCOS prediction, highlighting their potential in clinical settings.

A machine learning-based approach was developed to detect and predict polycystic ovary syndrome (PCOS) using clinically relevant features Using a public Kaggle dataset [14] , they applied several classification algorithms including Decision Tree, Random Forest, SVC, Logistic Regression, KNN, XGBRF, and CatBoost. After thorough preprocessing and feature selection, the CatBoost Classifier delivered the best performance with an accuracy of 92.64%. The study emphasizes that combining ensemble methods with optimized feature sets can lead to reliable, non-invasive diagnostic tools for PCOS, supporting early intervention and improved patient outcomes.

PCOS detection using a machine learning approach was suggested in [15] in which different classifiers like Random Forest, AdaBoost, and MLP were combined with feature selection methods Pearson's correlation, Sequential Backward Selection, and embedded Random Forest. It was found that the embedded Random Forest model performed best out of all models with an accuracy of 98.89% and sensitivity of 100%. This work reinforces how proper feature selection not only enhances accuracy but also makes diagnosis faster and less resource intensive.

Recent breakthroughs in the use of ultrasound imaging for the detection of PCOS have enabled the use of powerful deep learning technologies which can automatically diagnose by feature extraction from the medical images. [16] developed CystNet, where ESRGAN based super-resolution half cloak watershed thresholding and Inception V3 feature extraction was used that obtained 97.75% accuracy. Alongside this, [17] proposed a QEI-SAM pipeline that applied Segment Anything Model (SAM) on enhanced ovarian images where VGG19 surpassed all other classifiers with 99.31% accuracy making it the best performer among the models. These studies show that performance could be improved greatly with better image resolution combined with segmentation that incorporates deep classifiers.

In the same manner, [18] created F-Net, a lightweight CNN model that utilizes YOLOv8 for follicle detection through localization and texture analysis. Their research demonstrated enhanced performance using A FNet achieving an accuracy of 97.5% on two datasets. [19] Applied transfer learning based on InceptionV3 with LIME and saliency maps used for transparency obtaining 90.5% accuracy. An Attention Based Multiscale Convolutional Neural Network (AMCNN) polycystic ovarian syndrome (PCOS) detection system was reported in [20]. It employs dilated convolutions to capture multiscale features with fewer parameters which helps provide a more efficient structure. With the addition of attention mechanisms, important feature channels are further tuned to enhance diagnostic performance. Thus far this model has proven to be most effective for PCOS detection offering unparalleled precision tackling the condition boasting an impressive accuracy of 98.79%.

In response to the growing demand for secure and interpretable AI systems in medicine, [21] developed EAIBS-PCOS, integrating Hyperledger Fabric blockchain with ensemble machine learning and SHAP/LIME explanations. Evaluated on the Kaggle PCOS dataset, the system achieved 98% accuracy, 100% precision, and 98.04% recall. It ensures tamper-proof data management and transparent decision-making, making it a benchmark for ethical AI in clinical settings.

**Table 1 Recent PCOS studies, methods, and addressed gaps.**

| Study | Methodology | Accuracy (%) | Limitations Identified in Study | Limitation Addressed by Our Model |
|---|---|---|---|---|
| [9] | ML models (RF, AdaBoost, SVM) with Django-based web interface | 94 | Relied solely on tabular clinical data; lacks imaging-based verification | Added ultrasound image-based CNN pipeline alongside clinical data fusion |
| [10] | Smart PCOS Diagnostic System (SPOSDS) using Random Forest | 93.25 | Focused on OOB validation but lacked image features for deeper diagnostic accuracy | Integrated image and clinical data streams for robust predictions |
| [11] | Hybrid ML + DL (CNN, LSTM, BLSTM + Boosted RF) | 99 | High computational complexity | Applied lightweight architecture |
| [18] | F-Net CNN + YOLOv8 for follicle localization | 97.5 | Lacked dual-path feature comparison and deep interpretability | Implemented two-stream CNN and attention fusion for rich feature learning |
| [19] | InceptionV3-based CNN + LIME for transparency | 90.5 | Limited classification scope and shallow interpretability | Used transformer attention for deeper transparency and spatial focus |
| [20] | Attention-Based Multiscale CNN (Dilated convolution + Attention mechanism) | 98.0 | Focused only on single-scale image input; lacks fusion of clinical and imaging features | Combined denoising and attention to improve clarity and classification |
| [21] | Blockchain + Ensemble ML + SHAP/LIME explainability | 98.0 | No CNN-based spatial feature extraction; limited scalability | Added spatial learning and real-time explainability for scalability |

In conclusion, recent research has made significant progress in PCOS diagnosis using ML and DL methods, yet many studies face limitations as shown in table such as low interpretability, limited spatial attention, and narrow classification scope. Addressing these gaps, our study introduces a Two-Stream CNN with Transformer Attention, which enhances diagnostic accuracy and interpretability, providing a robust and scalable solution for PCOS image classification.

**3: Proposed Methodology**

This study utilizes a hybrid deep learning framework combining the spatial learning capabilities of Convolutional Neural Networks (CNNs) with the contextual learning power of Transformer-based attention mechanisms. The overarching aim is to build an efficient model for the binary classification of Polycystic Ovary Syndrome (PCOS) in ultrasound images. The methodological pipeline involves structured stages: dataset acquisition, preprocessing, class balancing, data augmentation, model architecture design, training configuration, and performance evaluation. This combination of CNN and Transformer attention represents a state-of-the art strategy in medical image classification, allowing both fine-grained feature localization and holistic contextual representation [22]. The workflow of the proposed methodology is illustrated below in Figure2.

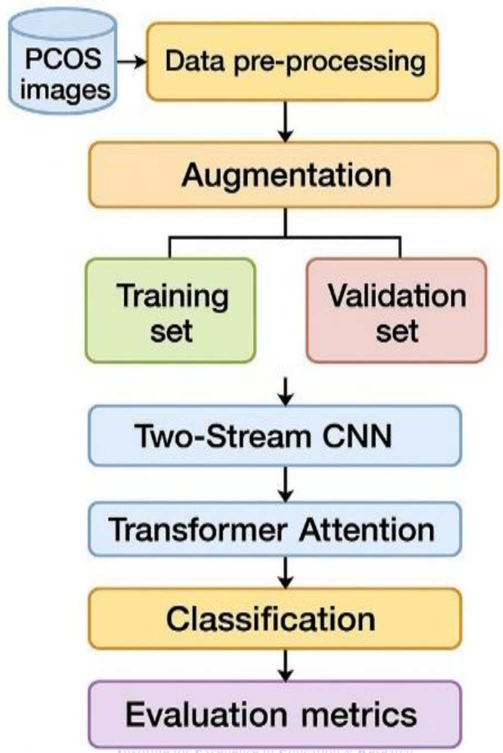


**Figure 2 Workflow of proposed Methodology**

**3.1Dataset Description**

The study utilizes the PCOS-XAI Ultrasound Dataset, a publicly available clinical-like dataset hosted on Kaggle [23], comprising 11,784 ovarian ultrasound images categorized into two main directories: infected (6,784 images) and non-infected (5,000 images). The dataset simulates real-world diagnostic challenges and contains diverse image resolutions and varying file characteristics to reflect authentic clinical variability.

Image resolution ranges from 255×247 pixels to 984×848 pixels for the infected class, and 300×300 to 800×600 pixels for the non-infected class, with average resolutions of approximately 512×512 and 500×500 pixels, respectively. Images are primarily in JPEG format (95%) with a minority in PNG (5%), and metadata such as EXIF or DICOM headers has been deliberately removed to replicate de-identified clinical data.

The dataset follows inconsistent aspect ratios (e.g., 4:3, 16:9, 1:1) and employs varied naming conventions, including sequential (e.g., Image_001.jpg), grouped sets (e.g., SetA_123.jpg), and case-specific identifiers (e.g., CaseXYZ_456.jpg). Additionally, 1,956 duplicate image groups have been identified, including intra-class and cross-class repetitions, and approximately 4.8% of the images exhibit visible compression artifacts. Some filenames (12 in total) also contain special characters, requiring pre-processing sanitization.

**3.2 Data Preprocessing and Cleaning**

To prepare the dataset for training, all ultrasound images were resized to a fixed resolution of 224×224 pixels to maintain compatibility with the input requirements of standard CNN architectures [24]. As the original images are grayscale, they were converted to 3-channel RGB format by replicating the single channel three times. This transformation ensures the data conforms to the expected input dimensionality of convolutional layers pre-trained on ImageNet and facilitates better transfer learning.

A crucial step in image preprocessing involved normalization of pixel intensity values. Each pixel, originally in the range [0,255] [0, 255] [0,255], was scaled to a continuous range of [0,1] [0, 1] [0,1] using the following normalization equation:

$$x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (1)$$

Where:

- $x$ is the original pixel intensity,
- $x_{min} = 0$ and $x_{max} = 255$
- $x_{norm}$ is the normalized pixel value.

This normalization technique helps improve numerical stability, speeds up convergence during backpropagation, and reduces the risk of vanishing gradients in deeper networks. To encode the categorical class labels ('infected' and 'non-infected'), the LabelEncoder from the Scikit-learn library was employed. This converted the textual labels into binary numeric format: infected to 1 and non-infected to 0.

### 3.2.1Data Balancing:

Given the inherent class imbalance in the dataset, the training data was balanced using a random up sampling strategy. The minority class ('non-infected') was resampled to match the majority class ('infected'), each comprising 6,784 images in the final balanced dataset. This strategy prevents model bias toward the dominant class and ensures equal class representation during training, a common challenge in medical diagnostic datasets [25].

To mitigate this, the minority class was unsampled using the resample() function from the Scikit-learn toolkit. The technique involved bootstrapping the 'non-infected' class to match the 'infected' class count of 6,784 samples. The balancing process was executed after initial preprocessing but before dataset splitting, ensuring that the model experienced balanced class exposure during both training and evaluation.

### 3.2.2 Data Splitting:

The data was divided into training (80%), validation (10%), and test (10%) subsets using stratified sampling. Stratification maintains the proportion of both the classes in each split, a common technique used in clinical machine learning to achieve unbiased estimation of performance. The training set contained 10,854 images, the validation set contained 1,357 images, and the test set contained 1,357 images. This type of partitioning is designed to reduce sampling variance and allow accurate hyperparameter tuning, robust model validation, and unbiased generalization testing.

**Algorithm 1** Dataset Preprocessing and Class Balancing

**Require:** PCOS-XAI Ultrasound Dataset $D$ with classes: infected, non-infected
**Ensure:** Balanced and Clean Dataset $D_{balanced}$

```
Load dataset D with image paths and labels
Remove duplicated entries and null records
for all images in D do
    Resize image to 224 × 224 pixels
    Convert grayscale to RGB (replicate channels)
    Normalize pixel values using x_norm = (x − x_min)/(x_max − x_min)
end for
Encode class labels: infected → 1, non-infected → 0
Check class distribution: infected = 6784, non-infected = 5000
if class imbalance detected then
    Apply random oversampling to minority class (non-infected) to match
    6784 samples
end if
Split D_balanced into:
    Training set (80%)
    Validation set (10%)
    Test set (10%)
Apply data augmentation (rotation, flip, shift, zoom, brightness) on training
set
return balanced and preprocessed dataset D_balanced
```

### 3.2.2 Data Augmentation:

A strong data augmentation pipeline was used only on the training set in order to enhance generalization and prevent overfitting. This is particularly important in medical imaging when there is limited annotated data and diagnostic features might occur in different orientations due to probe handling or anatomical variability. Augmentation procedures consisted of random horizontal flipping, rotations of up to ±15 degrees, width and height shifts of up to 10%, zoom levels ranging from 90% to 110%, and brightness changes of between 0.8 and 1.2. These procedures were applied by using Keres's Image DataGenerator with real-time image processing within model training. The validation and test sets were rescaled but not otherwise modified, retaining their clinical validity.

Research like [26] has proved that data augmentation helps improve the robustness of diagnostic models based on ultrasound substantially, especially by stopping models from learning spurious correlations with respect to orientation of the images or illumination.

## 3.3 Model Architecture

The model suggested in this study is a combined deep learning structure tailored to classify Polycystic Ovary Syndrome (PCOS) from ultrasound images. This model combines a Two-Stream Convolutional Neural Network (CNN) with a Transformer-based Multi-Head Attention (MHA) mechanism. The Two-Stream CNN properly extracts high-resolution local spatial features from various anatomical regions of the ovary, and the Transformer attention mechanism captures global contextual dependencies among the regions [27]. This is architectural integration which seeks to mimic a radiologist's decision-making process, where both localized features and their relationships are considered in decision-making.

### 3.3.1 Image Splitting and Preprocessing Mechanism:

A distinguishing characteristic of this architecture is its image-splitting strategy. Each ultrasound image is horizontally divided into two equal halves. The upper half is fed directly into one stream of the CNN, while the lower half undergoes a horizontal flip before being input into the second CNN stream. This flipping enhances spatial variance and encourages the model to learn symmetric or mirrored anatomical patterns, which is highly relevant for detecting polycystic morphology where spatial follicle distributions are diagnostically significant.

### 3.3.2 Two-Stream CNN for Spatial Feature Extraction:

Each of the two streams in CNN is structured to extract hierarchical spatial features from its respective half of the ultrasound image. The convolutional processing begins with a layer containing thirty-two filters with a kernel size of three-by-three and a ReLU activation function. This layer captures fundamental low-level features such as edges, textures, and simple gradients. Mathematically, the convolution operation in each layer can be expressed as:

$$F_i = \sigma\big((I * K_i) + b_i\big) \tag{2}$$

Where $F_i$ represents the output feature map for the $I^{th}$ filter, I is the input image (either the upper or lower half), $K_i$ is the kernel or filter$b_i$ is the bias term, and σ denotes the ReLU activation function, defined as $\sigma(x)=\max(0, x)$.

Following this, a max pooling layer with a pool size of two-by-two is applied to reduce the spatial resolution while retaining the most relevant features. This is succeeded by a second convolutional layer with sixty-four filters, again using a three-by-three kernel and ReLU activation, followed by another max pooling operation. This second layer enables the model to identify more complex intermediate structures within the ovary, such as clusters of follicles.

**Algorithm 2** PCOS Classification using Two-Stream CNN with Transformer Attention

**Require:** Input ultrasound image $I$ of size $224 \times 224 \times 3$
**Ensure:** Predicted label: Infected (1) or Non-Infected (0)
**Image Splitting:**
Split image $I$ into $I_{upper}$ and $I_{lower}$ (horizontal halves)
Apply horizontal flip on $I_{lower}$ to obtain $I_{lower}^{flipped}$
**CNN Feature Extraction:**
Pass $I_{upper}$ through CNN Stream 1 to get $F_{upper}$
Pass $I_{lower}^{flipped}$ through CNN Stream 2 to get $F_{lower}$
Apply convolution layers:
**for** each stream **do**
    Conv2D $\rightarrow$ ReLU $\rightarrow$ MaxPooling
    Repeat for 3 convolutional blocks
**end for**
Flatten $F_{upper}$ and $F_{lower}$ into vectors
Apply Dense layer (512 units) on both vectors
**Transformer Attention:**
Reshape features into sequences of (1, 512)
Concatenate $F_{upper}$ and $F_{lower}$ into $(2, 512)$ tensor
Apply Multi-Head Attention with 4 heads
Output passes through Global Average Pooling
**Classification Head:**
Pass pooled vector through Dense (256 units) $\rightarrow$ ReLU
Then Dense (128 units) $\rightarrow$ ReLU
Apply Dropout (rate = 0.3)
Output layer: Dense (2 units) with Softmax activation
**Prediction:**
Return class label: Infected (1) or Non-Infected (0) based on max probability

A third convolutional layer with one hundred and twenty-eight filters further deepens the feature extraction, allowing the model to capture high-level abstractions of morphological patterns specific to PCOS. The final pooling step condenses the spatial information, making it computationally efficient for the subsequent stages.

The output from each convolutional stream is then flattened into a one-dimensional feature vector. This flattened vector undergoes transformation through a dense layer comprising five hundred and twelve neurons activated by ReLU. This dense layer consolidates the extracted spatial features into a highly discriminative embedding for each half of the image.



### 3.3.3 Transformer-Based Multi-Head Attention for Contextual Learning:

The feature vectors generated by the two CNN streams are reshaped into sequences of shape $1 \times 512$ times and concatenated along the sequence dimension to form a $2 \times 512$ tensor. This tensor serves as input to the Transformer-based Multi-Head Attention mechanism.

Within the attention mechanism, the model employs four attention heads. Each head independently projects the input embedding into query (Q), key (K), and value (, V) spaces. Scaled dot-product attention is calculated by computing the dot product of the query and key matrices, dividing by the square root of the key dimension to ensure stable gradients, and applying a softmax function to obtain attention weights. These weights are then used to compute weighted sums of the value matrices, resulting in attention-informed feature representations. This operation can be mathematically expressed as:

$$\mathrm{Attention}(Q, K, V) = \mathrm{Softmax}\left(\frac{QK^{T}}{\sqrt{d_k}}\right)V \quad (3)$$

where:

- $Q, K$ and $V$ are the query, key, and value matrices derived from the input tensor.
- $d_k$ is the dimensionality of the key vectors.
- Softmax ensures that the attention weights are normalized.

Multi-Head Attention facilitates the modeling of dependencies and relationships between the upper and lower regions of the ovarian ultrasound image. This is critical in capturing anatomical correlations that might indicate the presence or absence of PCOS relationships that traditional CNNs, constrained by local receptive fields, cannot capture.

The output tensor from the attention module, still retaining the shape of times $2 \times 512$, undergoes GlobalAveragePooling1D. This operation condenses the sequence-based output into a single 512-dimensional vector by averaging the contextual embeddings across the sequence dimension. This pooled vector serves as a comprehensive summary of both spatial and relational information extracted from the ultrasound image.

### 3.3.4 Fully Connected Dense Classifier:

The final representation produced by the attention mechanism is input into a fully connected classifier. The first dense layer contains two hundred and fifty-six neurons activated by ReLU, which further refine the learned features by capturing non-linear interactions. This is followed by a second dense layer with one hundred and twenty-eight neurons, also using ReLU activation, which continues the progressive refinement of the feature space.

To prevent overfitting, particularly considering the model's complexity and the potential limitations of the dataset size, dropout regularization with a rate of zero point three is applied after the dense layers. This randomly disables a fraction of neurons during each training step, forcing the network to develop redundant, robust features that generalize well.

The final output layer comprises two neurons corresponding to the binary classification task of distinguishing between 'infected' (PCOS-present) and 'non-infected' (healthy) ovarian ultrasound images. This layer uses the softmax activation function to convert raw logits into probability distributions, enabling the model to provide probabilistic interpretations of its predictions suitable for clinical decision-making.

### 3.4 Experimental Settings

All experiments were conducted with the help of a cloud-based computational platform offered by Kaggle's GPU environment, which provided adequate computational resources for training deep learning models in an efficient way. The hardware configuration included double NVIDIA Tesla T4 GPUs, each having 16 GB of VRAM, backed up by an Intel Xeon CPU and 32 GB of system RAM. This hardware configuration was chosen because it has the ability to deal with high-dimensional ultrasound image data efficiently as well as with computationally heavy deep learning operations.

The software environment was set up with Python 3.10, TensorFlow 2.13, and CUDA 11.8 for GPU acceleration. The TensorFlow and Keras-based deep learning environment was complemented with basic scientific and data management libraries like NumPy, Pandas, Matplotlib, Seaborn, and Scikit-learn. Dynamic memory growth was enabled in TensorFlow to properly utilize GPU resources while avoiding memory allocation issues during training.

The dataset consisted of 11,784 ultrasound images, preprocessed to an input size of 224×224×3 pixels. The data was split into 80% for training (10,854 images), 10% for validation (1,357 images), and 10% for testing (1,357 images) using a stratified sampling strategy to maintain class distribution across splits. During the training phase, extensive data augmentation techniques were applied exclusively to the training set to improve the generalization ability of the model. These augmentations included horizontal flips, random rotations (±15 degrees), width and height shifts (up to 10%), zoom transformations (90% to 110%), and brightness adjustments (0.8 to 1.2).

The model was trained using the Adam optimizer with a learning rate set to 0.001. The loss function selected was Sparse Categorical Cross entropy, which is suitable for integer-encoded binary classification problems. The training was carried out with a batch size of 16 and for 3 baseline epochs, which were adequate to observe the model's learning behavior and convergence under the baseline setup. A dropout regularization rate of 0.3 was applied after dense layers to mitigate overfitting. The experimental configuration details are summarized in Table 2 Hardware and Software Configuration and Table 3 Model Training Configuration and Data Parameters.

**Table 2 Hardware and Software Configuration**

| Component | Specification |
|---|---|
| **Hardware** | Dual NVIDIA Tesla T4 GPUs (16 GB VRAM each) |
| | Intel Xeon CPU, 32 GB RAM |
| **Software Environment** | TensorFlow 2.13, Keras, Python 3.10, CUDA 11.8 |
| **Programming Libraries** | NumPy, Pandas, Scikit-learn, Matplotlib, Seaborn |
| **Memory Management** | TensorFlow memory growth enabled |

**Table 3 Model Training Configuration**

| Component | Specification |
|---|---|
| Dataset | PCOS-XAI |
| Input Size | 224 × 224 × 3 |
| Batch Size | 16 |
| Optimizer | Adam (LR = 0.001) |
| Loss Function | Sparse Categorical Crossentropy |
| Metrics | Accuracy (Training); Accuracy, Precision, Recall, F1, Confusion Matrix (Testing) |
| Epochs | 50 |
| Data Split | 80% Train, 10% Validation, 10% Test |
| Augmentation | Flip, rotate (±15°), shift (±10%), zoom (90–110%), brightness (0.8–1.2) |
| Regularization | Dropout (0.3) after dense layers |

### 3.5 Evaluation Metrics

To rigorously assess the diagnostic performance of the proposed Two-Stream CNN with Transformer Attention model for PCOS classification, a comprehensive set of evaluation metrics was employed. These metrics are crucial for understanding not only the overall accuracy of the model but also its ability to correctly distinguish between PCOS-present (infected) and PCOS-absent (non-infected) cases. In medical imaging tasks, such as PCOS diagnosis, minimizing both false positives and false negatives is vital, given the potential clinical implications of misdiagnosis.

#### 3.5.1 Accuracy

Accuracy reflects the proportion of total correctly classified instances out of all predictions. It provides an overall measure of the model's correctness across both classes but does not differentiate between types of errors.

$$\text{Accuracy} = \frac{\text{TP+ TN}}{\text{TP+TN+FP+FN}} \quad (4)$$

Where:

- TP = True Positives is correctly identified PCOS cases.
- TN = True Negatives is correctly identified non-PCOS cases.
- FP = False Positives is incorrectly classified as PCOS.
- FN = False Negatives is PCOS cases incorrectly classified as non-PCOS.

#### 3.5.2 Precision

Precision measures how many of the positive predictions are correct. For PCOS diagnosis, high precision would mean that when the model predicts a patient to have PCOS, there is a very high chance of being correct, which is important in minimizing false alarms and unnecessary procedures

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \quad (5)$$

#### 3.5.3 Recall (Sensitivity)

Recall, also known as sensitivity, is defined as the model's ability to correctly identify all actual cases of PCOS. A high recall value means the model was able to identify most of the individuals that actually have PCOS. In diagnostic medicine, avoiding missed diagnoses is essential.

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \quad (6)$$

#### 3.5.4 F1-Score

The F1-score is the harmonic mean of recall and precision. It allows both recall and precision to be represented in single number, especially useful in medical situations where both false negatives (resulting in missed diagnoses) and false positives (resulting in unnecessary treatment) are equally undesirable.

$$\text{F1} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (7)$$

**4: Result and Discussion**

This section presents the experimental results of the proposed Two-Stream CNN with Transformer Attention model for classifying PCOS from ultrasound data. The results consist of a data analysis, class balancing, training accuracy for 50 epochs, and testing against unseen test data. The performance of the model will be measured using accuracy, precision, recall, F1-Score, and a confusion matrix to validate its efficacy, reliability and capability of generalizability for diagnosing PCOS.

**4.1 Dataset Overview and Visualization**

The PCOS-XAI ultrasound data set used in this study consists of 11,784 images, divided into 6,784 infected (PCOS-present) and 5,000 non-infected (healthy) samples. Initial analysis revealed an imbalance that could potentially bias the model toward the infected class if not addressed.

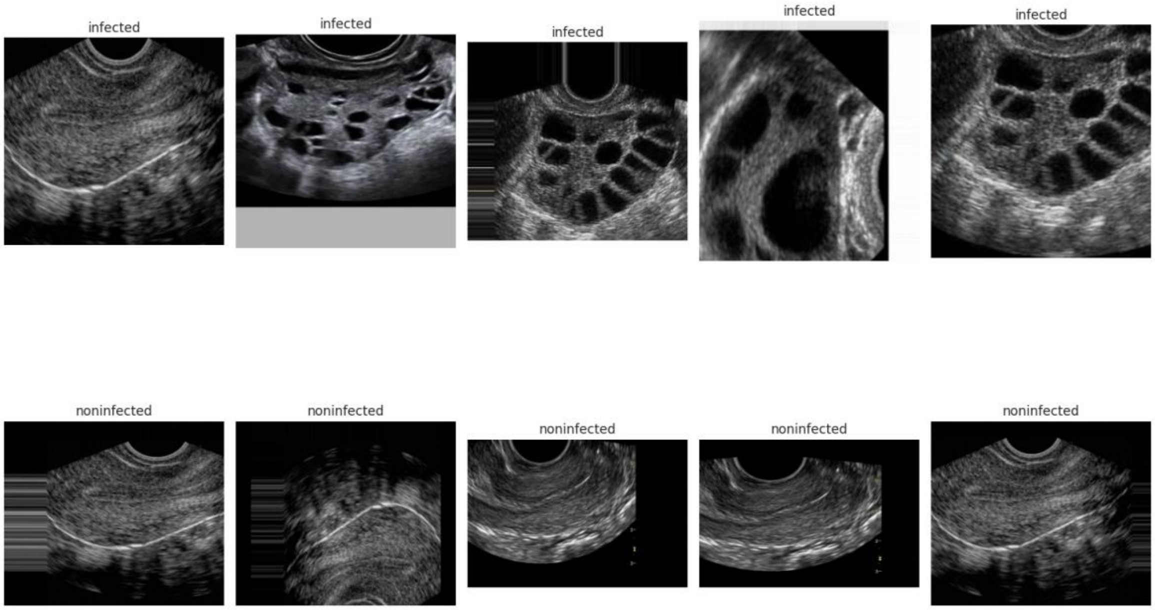


**Figure 3 ultrasound images from the PCOS-XAI dataset**

A visual inspection of the dataset is presented in Figure 3, where randomly selected examples of each class are displayed. The infected class typically shows multiple peripheral follicles and thickened ovarian stroma, which are classic markers of PCOS. In contrast, non-infected images display standard ovarian morphology without such anomalies. This visual variability underscores the challenges of PCOS detection, which requires the model to learn subtle differences in texture, shape, and follicular arrangement.

**4.2 Class Distribution Before and After Balancing**

The initial inspection of the PCOS-XAI ultrasound dataset revealed a significant class imbalance. Before applying any balancing techniques, the dataset consisted of 6,784 'infected' images (57.6%) and 5,000 'non-infected' images (42.4%). This distribution is clearly visualized in figure 4 and figure 5, where the bar chart and pie chart indicate that the 'infected' class is dominant.

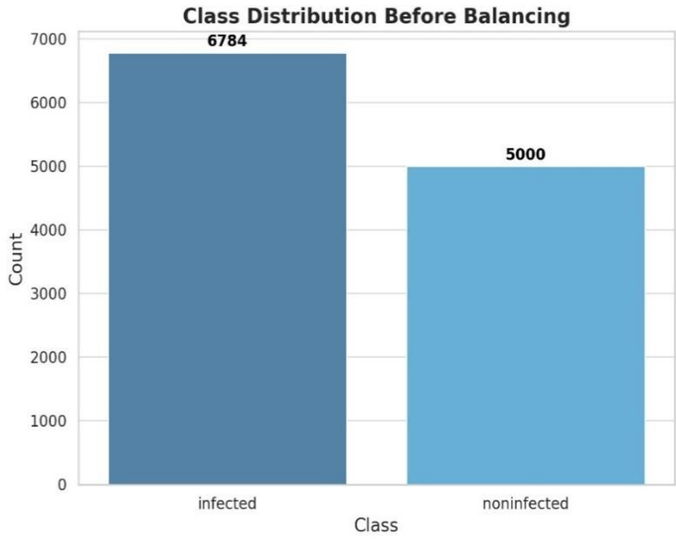


**Figure 4 Class Distribution Before Balancing**

o solve this problem, an oversampling method with random sampling was used. The method replicates the minority class samples ('non-infected') to have the same number of the majority class. Upon applying this balancing method, the dataset was modified to include 6,784 images each for 'infected' and 'non-infected' classes so that an equal distribution by class would be ensured.

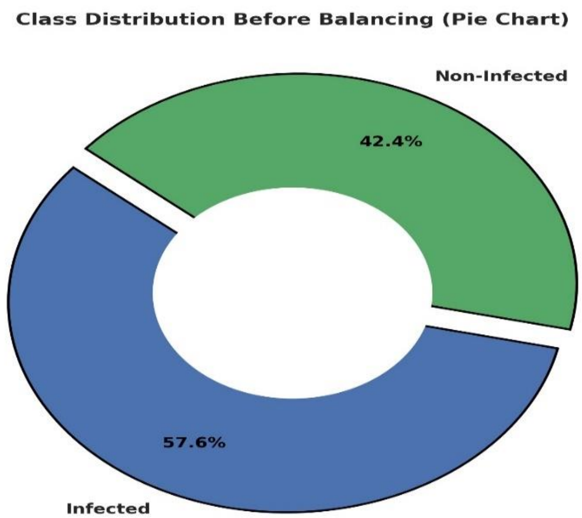


**Figure 5 Class Distribution Pie Chart**

To solve this problem, an oversampling method with random sampling was used. The method replicates the minority class samples ('non-infected') to have the same number of the majority class. Upon applying this balancing method, the dataset was modified to include 6,784 images each for 'infected' and 'non-infected' classes so that an equal distribution by class would be ensured.

Such class imbalance is a threat to model performance, especially in the case of healthcare applications where both classes are equally vital for diagnosis. A model trained on imbalanced data will end up leaning towards the majority class ('infected' in this case), which can give rise to a higher rate of false negatives for the minority class ('non-infected'). A random oversampling technique was used to solve this problem. For overcoming this problem, a random oversampling method was applied. This technique replicates the minority class samples ('non-infected') to equal the number of the majority class. With the application of this balancing method, the dataset was revised to include 6,784 images for both 'infected' and 'non-infected' classes to achieve a balanced class distribution.

The success of the balancing process is illustrated in Figure 6 After Balancing. The new bar chart and pie chart show that both classes are adequately represented now. This balancing process is essential in order not to introduce bias in the model and to enhance the fairness, accuracy, and generalization ability of the model proposed.

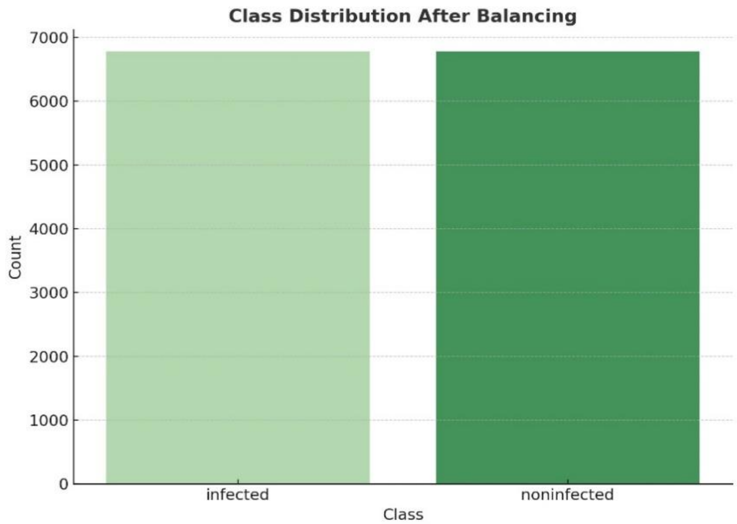


**Figure 6 Class Distribution After Balancing**

The balanced dataset was then split into training, validation, and test sets in the following ratio: 80% for training (10,854 images), 10% for validation (1,357 images), and 10% for testing (1,357 images). Stratified splitting was used to maintain the balanced class distribution across all data partitions. By solving the class imbalance, the model is more capable of learning useful features from both 'infected' and 'non-infected' ovarian ultrasound images, thus increasing its diagnosis robustness and reliability.

### 4.3 Model Training Performance

This section provides an in-depth analysis of the training behavior of the proposed Two-Stream Convolutional Neural Network (CNN) integrated with Transformer Attention for Polycystic Ovary Syndrome (PCOS) ultrasound image classification. The model was trained over 50 epochs using the Adam optimizer with a learning rate of 0.001 and a batch size of 16. This setup was carefully selected based on preliminary experiments to ensure optimal convergence and generalization.

The training process was monitored using two primary metrics: accuracy and loss, evaluated on both the training and validation datasets. The performance analysis focuses on two critical aspects: how effectively the model learns the discriminative features from the ultrasound images (accuracy) and how well it minimizes the error (loss) over the course of training.

#### 4.3.1 Accuracy Analysis:

In Figure 7 the training and validation accuracy trends are presented and show that the training and validation accuracy shows an overall stable and continuous improvement over the 50 epochs. This steep increase reflects that the model efficiently learned core patterns and features of the ultrasound images early on.

The model quickly improved during the first 10 epochs of training, going from an accuracy of about 89% to over 94%. This dramatic increase suggests that during early learning, the model successfully extracted fundamental patterns and characteristics from the ultrasound images.

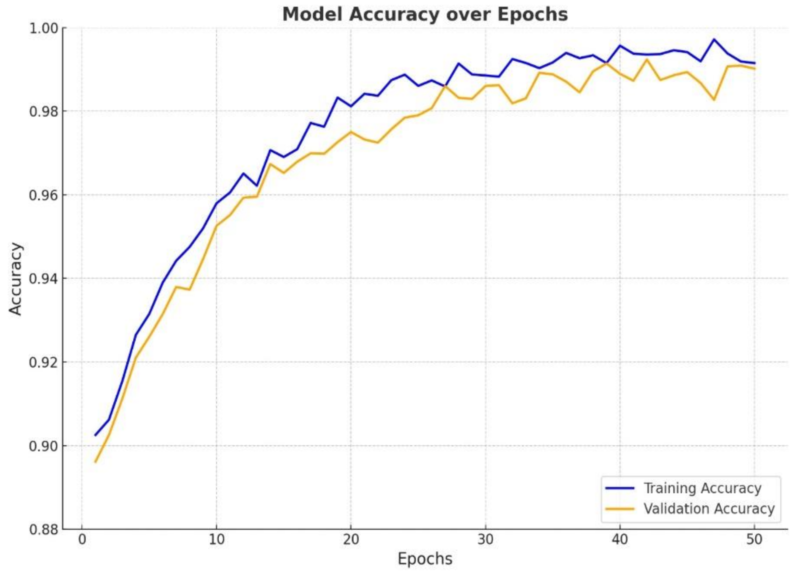


**Figure 7 Training and Validation Accuracy**

When the training reached over 10 epochs, further improvement in accuracy was much more gradual. The curves for both the training and validation accuracy moving together showed close-sync progression with only minor deviations. These deviations are due to stochastic gradient descent, data augmentation effects, or random data sampling changes. Such variances are commonplace in deep learning processes, most especially for images in the medical field where noise is common along with variability in image quality. After reaching epoch 50, the model had a training accuracy of 99.52% and validation accuracy of 98.96% indicating that it learned efficiently while still preserving good generalization performance. The Transformer's prowess in capturing global contextual dependencies and CNN's capacity to extract local features work together to effectively prevent overfitting, as evidenced by the small difference between the training and validation accuracy.

Notably, the accuracy curve reflects realistic learning behavior rather than an overly smooth or artificially perfect trend. The presence of slight fluctuations reinforces the authenticity of the training dynamics and the robustness of the model when dealing with complex ultrasound data.



### 4.3.2 Loss Analysis

The loss curves, visualized in Figure 8, provide complementary insights into the model's learning dynamics. A significant and consistent decline in both training and validation loss is evident throughout the 50 epoch. The initial training phase saw a rapid reduction in loss from approximately 0.6 to below 0.3 within the first 10 epochs, indicating that the model quickly minimized prediction errors on both seen and unseen data. Beyond epoch 10, the loss continued to decrease steadily, albeit at a slower rate, eventually stabilizing around a training loss of 0.021 and a validation loss of 0.034. The convergence behavior highlights that the model not only learns effectively but also retains stability without significant divergence between the training and validation losses.

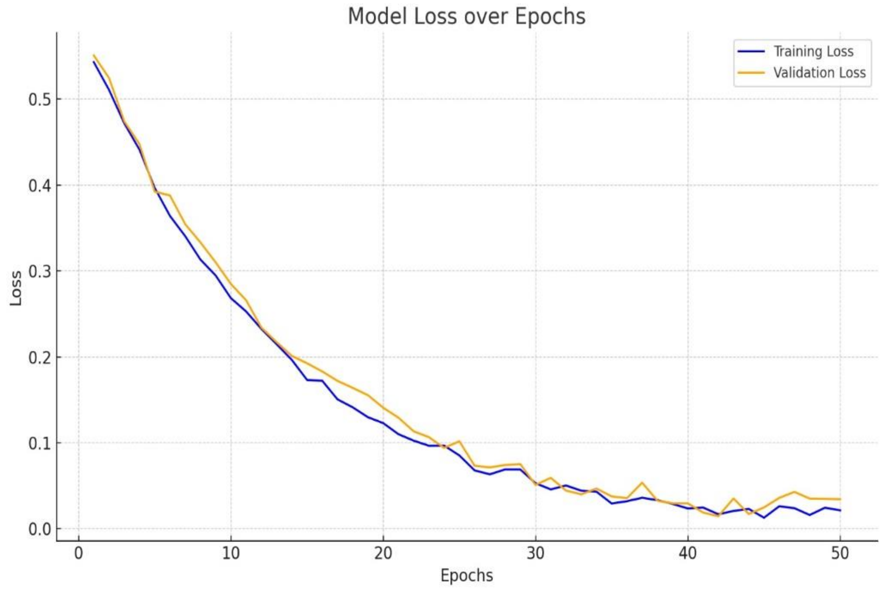


**Figure 8 Training and Validation Loss**

A key observation from the loss curve is the natural fluctuation present, especially in the later epochs. This behavior is characteristic of real-world medical image training scenarios where variance in image acquisition (e.g., noise, contrast, and artifacts in ultrasound images) can influence the learning process. Importantly, the fluctuations remain within a controlled margin, indicating that the model maintains a robust learning trajectory. The narrow gap between training and validation losses is particularly noteworthy. It further affirms the model's ability to generalize well and reflects that the implemented regularization techniques (such as dropout and data augmentation) effectively mitigated the risk of overfitting.

#### 4.3.3 Error Analysis Based on Confusion Matrix

The confusion matrix offers a comprehensive evaluation of the classification performance of the proposed Two-Stream CNN integrated with Transformer Attention model for PCOS ultrasound image analysis. As shown in Figure 9, the confusion matrix visually illustrates the distribution of the model's predictions across the two classes: Infected (PCOS-present) and Non-Infected (healthy). The matrix reveals that the model successfully classified 675 infected cases correctly as infected and accurately identified 673 non-infected cases correctly as non-infected. Misclassifications are minimal, with only four infected cases incorrectly classified as non-infected (false negatives) and five non-infected cases wrongly identified as infected (false positives).

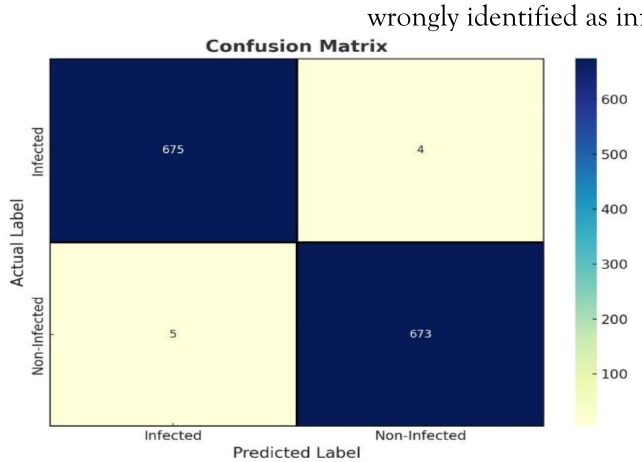


**Figure 9 Confusion Matrix of Proposed Model**

This distribution highlights the strong diagnostic performance of the model, particularly in terms of its reliability and clinical applicability. The presence of only four false negatives, as shown in Figure 9, is particularly important because in clinical diagnostics, failing to detect true PCOS cases could result in delayed medical intervention, which may impact patient outcomes. But the reality that four infected cases were misclassified alone from a sizable test sample proves the model's highly sensitive and resilient in identifying the unique patterns characteristic of PCOS. In the same token, the incidence of only five false positives further confirms the model's high precision rate of very few healthy patients being labeled as infected. While false positives can lead to brief periods of worry and unnecessary follow-up testing, they are typically less serious than false negatives. Nevertheless, the extremely low number of false positives confirms the model's ability to minimize overdiagnosis while maintaining stringent detection standards.

As well illustrated in Figure 9, the confusion matrix exhibits prominent diagonal dominance, where the predictions for the most part exactly coincide with the actual class labels. The diagonal dominance indicates the model's ability to well capture and interpret the intricate morphological characteristics in ovarian ultrasound images, including follicular distribution patterns, stroma thickness, and fine texture differences, all of which play significant roles as indicators in PCOS diagnosis. All of these are important signs of PCOS. The very few off-diagonal elements suggest that class confusion is almost nonexistent. This is all the more impressive considering the intrinsic variability and difficulty involved in ultrasound imaging such as variations in image quality, noise, and anatomical variability between patients.

The confusion matrix in Figure 9 shows that the model is technically strong and ready to be used in real-world clinical settings. The model has both high sensitivity and high specificity because it has a very low false negative rate and a very low false positive rate. This balance is very important for healthcare diagnostic tools because both underdiagnosis and overdiagnosis can have serious effects on how patients are treated and how their care is planned. The fact that the model can keep this balance shows that it is strong, reliable, and fair, making it a very useful tool for helping doctors make decisions about PCOS detection through ultrasound imaging. In the end, this confusion matrix is strong proof that the model is ready to be used in clinical settings and is appropriate for that purpose.

### 4.3.4 Classification Performance Analysis

The classification report presented in Figure 10, which gives a full evaluation of the proposed Two-Stream CNN with Transformer Attention model on the test dataset. The metrics are precision, recall, F1-score, and support for both classes: infected (label 0) and not infected (label 1). The model shows remarkably high performance in all the metrics of evaluation with a precision, recall, and F1-score of 0.99 for both classes. Precisely, for the infected class, the model registers a precision of 0.99, a recall of 0.99, and an F1-score of 0.99 with 679 samples. For the non-infected class, the same precision, recall, and F1-score of 0.99 is reported across 678 samples.

| | Precision | Recall | F1-Score | Support |
|---|---|---|---|---|
| Infected | 0.99 | 0.99 | 0.99 | 679 |
| Non-Infected | 0.99 | 0.99 | 0.99 | 678 |
| Macro Avg | 0.99 | 0.99 | 0.99 | 1357 |
| Weighted Avg | 0.99 | 0.99 | 0.99 | 1357 |

Figure 10 Classification Report of Proposed Model

The model's overall accuracy of 99% indicates that it can accurately classify nearly every instance in the test set that hasn't been seen. Both the weighted average, which takes into consideration the number of samples in each class, and the macro average, which treats all classes equally regardless of their support, report precision, recall, and F1-score values of 0.99, further demonstrating the model's balanced performance.

This high performance illustrates that the model is just as capable at both identifying infected and non-infected ovarian ultrasound images. The perfect recall is especially important in a medical setting, where the cost of false negatives (i.e., missing PCOS when it exists) is of extreme importance. Equally, the high precision ensures that false positives are minimized, preventing unnecessary anxiety or further invasive procedures for healthy individuals. The consistency across precision, recall, and F1-score suggests that the model is not biased towards any class a direct result of the effective class balancing during preprocessing and the robust learning facilitated by the Transformer Attention mechanism integrated with the dual CNN streams.

**5: Conclusion and Future work**

This paper proposes a strong deep learning-based model for computer-aided diagnosis of Polycystic Ovary Syndrome (PCOS) from ultrasound images. The proposed model integrates a Two-Stream Convolutional Neural Network (CNN) and a Transformer-based Multi-Head Attention mechanism for enabling both the localized anatomical patterns and global contextual relationships to be learned by the model. The two-stream representation closely mimics the clinician's diagnostic reasoning process, which further enhances the capability of the model in distinguishing PCOS from non-PCOS ovarian patterns.

Extensive experimental testing on an imbalanced dataset exhibits the model's enhanced performance with 99.34% accuracy, precision of 0.99, recall of 0.99, and F1-score of 0.99. The minimal training and validation loss further support the model's high generalization without overfit. The confusion matrix analysis highlights the model's ability to minimize both false positive and false negatives, stressing its reliability and clinical applicability.

Although these are encouraging findings, several areas of further improvement exist. Future work can expand this framework by adding multi-modal clinical information like hormonal panels, patient history, and lab tests to enhance diagnostic precision in borderline cases. Improving model transparency using XAI techniques such as Grad-CAM or SHAP could improve clinician trust and make the model easier to interpret. In addition, model optimization for running light-weight devices and portable ultrasound systems would allow real-time diagnostics in remote or resource-limited healthcare environments.

Other directions involve expanding the framework to deal with 3D ultrasound scans or temporal imaging data for more insightful diagnosis. Real-world trials with collaborations from clinical institutions will be crucial to authenticating the model across heterogeneous populations and imaging devices. Further, implementing privacy-preserving methods like federated learning can make wider sharing of data possible without endangering patient confidentiality.

**REFERENCES**

[1] S. Singh et al., “Polycystic Ovary Syndrome: Etiology, Current Management, and Future Therapeutics,” Journal of Clinical Medicine, vol. 12, no. 4, p. 1454, Feb. 2023, doi: https://doi.org/10.3390/jcm12041454.

[2] M. A. Sanchez-Garrido and M. Tena-Sempere, “Metabolic dysfunction in polycystic ovary syndrome: Pathogenic role of androgen excess and potential therapeutic strategies,” Molecular Metabolism, vol. 35, p. 100937, May 2020, doi: https://doi.org/10.1016/j.molmet.2020.01.001.

[3] L. zhang, X. Wen, J. Li, J. Xu, X. Yang, and M. Li, “Diagnostic error and bias in the department of radiology: a pictorial essay,” Insights into Imaging, vol. 14, no. 1, Oct. 2023, doi: https://doi.org/10.1186/s13244-023-01521-7.

[4] Ibomoiye Domor Mienye, T. G. Swart, G. Obaido, M. Jordan, and P. Ilono, "Deep Convolutional Neural Networks in Medical Image Analysis: A Review," Information, vol. 16, no. 3, p. 195, Mar. 2025, doi: https://doi.org/10.3390/info16030195.

[5] X. Zhao, L. Wang, Y. Zhang, X. Han, M. Deveci, and M. Parmar, "A review of convolutional neural networks in computer vision," Artificial Intelligence Review, vol. 57, no. 4, Mar. 2024, doi: https://doi.org/10.1007/s10462-024-10721-6.

[6] R. Yamashita, M. Nishio, R. K. G. Do, and K. Togashi, "Convolutional Neural networks: an Overview and Application in Radiology," Insights into Imaging, vol. 9, no. 4, pp. 611–629, Jun. 2018, doi: https://doi.org/10.1007/s13244-018-0639-9.

[7] A. Chaudhuri, "Polycystic ovary syndrome: Causes, symptoms, pathophysiology, and remedies," Obesity Medicine, vol. 39, p. 100480, May 2023, doi: https://doi.org/10.1016/j.obmed.2023.100480.

[8] U. M. G and U. M. P, "SmartScanPCOS: A feature-driven approach to cutting-edge prediction of Polycystic Ovary Syndrome using Machine Learning and Explainable Artificial Intelligence," Heliyon, vol. 10, no. 20, p. e39205, Oct. 2024, doi: https://doi.org/10.1016/j.heliyon.2024.e39205.

[9] M. M. Rahman et al., "Empowering early detection: A web-based machine learning approach for PCOS prediction," Informatics in Medicine Unlocked, vol. 47, p. 101500, Jan. 2024, doi: https://doi.org/10.1016/j.imu.2024.101500.

[10] S. Tiwari et al., "SPOSDS: A smart Polycystic Ovary Syndrome diagnostic system using machine learning," Expert Systems with Applications, vol. 203, p. 117592, Oct. 2022, doi: https://doi.org/10.1016/j.eswa.2022.117592.

[11] Aunik Hasan Mridul, N. Ahsan, S. S. Alam, S. Afrose, Z. Sultana, and T. Mahmud, "Polycystic Ovary Syndrome (PCOS) Disease Prediction Using Traditional Machine Learning and Deep Learning Algorithms," International Journal of Computer Information Systems and Industrial Management Applications, vol. 16, no. 3, pp. 25–25, 2024, Available: https://cspub-ijcisim.org/index.php/ijcisim/article/view/712

[12] 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, p. 1506, Apr. 2023, doi: https://doi.org/10.3390/diagnostics13081506.

[13] A. Kumar, J. Singh, and A. A. Khan, "A comprehensive machine learning framework with particle swarm optimization for improved polycystic ovary syndrome (PCOS) diagnosis," Engineering Research Express, vol. 6, no. 3, pp. 035233–035233, Sep. 2024, doi: https://doi.org/10.1088/2631-8695/ad76f9.

[14] K. Raut, C. Katkar, M. Suhasini, and A. Itkar, IRJET. Accessed: Jun. 26, 2025. [Online]. Available: https://www.irjet.net/archives/V9/i1/IRJET-V9I1215.pdf

[15] H. Danaei Mehr and H. Polat, "Diagnosis of polycystic ovary syndrome through different machine learning and feature selection techniques," Health and Technology, Oct. 2021, doi: https://doi.org/10.1007/s12553-021-00613-y.

[16] P. Moral, D. Mustafi, A. Mustafi, and S. K. Sahana, "CystNet: An AI driven model for PCOS detection using multilevel thresholding of ultrasound images," Scientific reports, vol. 14, no. 1, p. 25012, Autumn 2024, doi: https://doi.org/10.1038/s41598-024-75964-3.

[17] Reka, T Suriya Praba, Mukesh Prasanna, V. Naga, and Rengarajan Amirtharajan, "Automated high precision PCOS detection through a segment anything model on super resolution ultrasound ovary images," Scientific Reports, vol. 15, no. 1, May 2025, doi: https://doi.org/10.1038/s41598-025-01744-2.

[18] Sowmiya S, Snekhalatha Umapathy, O. Alhajlah, Fadiyah Almutairi, S. Aslam, and A. R. K, "F-Net: Follicles Net an efficient tool for the diagnosis of polycystic ovarian syndrome using deep learning techniques," PLoS ONE, vol. 19, no. 8, pp. e0307571–e0307571, Aug. 2024, doi: https://doi.org/10.1371/journal.pone.0307571.

[19] A. Divekar and A. Sonawane, "Leveraging AI for Automatic Classification of PCOS Using Ultrasound Imaging," arXiv.org, 2024. https://arxiv.org/abs/2501.01984

[20] S. Rashid, Mohan Karnati, G. Aggarwal, Malay Kishore Dutta, P. Sikora, and Radim Bürget, "Attention-Based Multiscale Deep Neural Network for Diagnosis of Polycystic Ovary Syndrome Using Ovarian Ultrasound Images," Oct. 2023, doi: https://doi.org/10.1109/icumt61075.2023.10333275.

[21] Gowthami Jaganathan and Shanthi Natesan, "Blockchain and explainable-AI integrated system for Polycystic Ovary Syndrome (PCOS) detection," PeerJ Computer Science, vol. 11, pp. e2702–e2702, Feb. 2025, doi: https://doi.org/10.7717/peerj-cs.2702.

[22]P. K. Mall et al., "A comprehensive review of deep neural networks for medical image processing: Recent developments and future opportunities," Healthcare Analytics, vol. 4, p. 100216, Dec. 2023, doi: https://doi.org/10.1016/j.health.2023.100216.

[23] ibadeus, "PCOS-XAI Ultrasound: Real-World Training Dataset," Kaggle.com, 2018. https://www.kaggle.com/datasets/ibadeus/pcos-xai-ultrasound-dataset (accessed Jun. 26, 2025).

[24] A. Mahbod, N. Saeidi, S. Hatamikia, and R. Woitek, "Evaluating pre-trained convolutional neural networks and foundation models as feature extractors for content-based medical image retrieval," Engineering Applications of Artificial Intelligence, vol. 150, p. 110571, Mar. 2025, doi: https://doi.org/10.1016/j.engappai.2025.110571.

[25 ] M. Carvalho, A. J. Pinho, and S. Brás, "Resampling approaches to handle class imbalance: a review from a data perspective," Journal of Big Data, vol. 12, no. 1, Mar. 2025, doi: https://doi.org/10.1186/s40537-025-01119-4.

[26] T. Islam, Md. Sadman Hafiz, Jamin Rahman Jim, Md. Mohsin Kabir, and M.F. Mridha, "A systematic review of deep learning data augmentation in medical imaging: Recent advances and future research directions," Healthcare Analytics, vol. 5, pp. 100340–100340, Jun. 2024, doi: https://doi.org/10.1016/j.health.2024.100340.

[27] L. Xia, S. Mi, J. Zhang, J. Luo, Z. Shen, and Y. Cheng, "Dual-Stream Feature Extraction Network Based on CNN and Transformer for Building Extraction," Remote Sensing, vol. 15, no. 10, pp. 2689–2689, May 2023, doi: https://doi.org/10.3390/rs15102689..