International Research Journal of Multidisciplinary Scope (IRJMS), 2025; 6(3):1123-1140

Original Article | ISSN (O): 2582-631X DOI: 10.47857/irjms.2025.v06i03.04808

SE-PCOSNet: Channel-Wise Attention Mechanisms for Automated PCOS Detection in Ultrasound Imaging

Chiranjib Dutta[1]*, Ananjan Maiti[2], Sandip Roy[3]

[1]Department of Computer Applications, Guru Nanak Institute of Technology, Kolkata, India, [2]Department of Computer Science and Engineering, Guru Nanak Institute of Technology, Kolkata, India, [3]Department of Computer Science and Engineering, JIS University, Kolkata, India. *Corresponding Author's Email: duttachiranjib@gmail.com

Abstract

This work presents SE-PCOSNet, a novel deep learning PCOS diagnosis model driven by ultrasound image analysis. The architecture extends the fundamental Convolutional Neural Network (CNN) with Squeeze-and-Excitation (SE) blocks that support dynamic feature channel recalibration. The addition of the SE blocks supports the model's ability to detect weak and faint features in gynecological ultrasound images, crucial to successful PCOS diagnosis. The suggested architecture consists of three convolutional blocks, each of which includes SE blocks for facilitating improved discriminative feature extraction. Apart from architectural enhancement, the technique utilizes state-of-the-art data preprocessing, data augmentation, and attention mechanisms. GPU acceleration, regularization techniques, and cross-validation enhance the training process to render it robust and avoid over fitting. With a vast and heterogeneous dataset of over 1,900 pelvic ultrasound images, SE-PCOSNet achieved impressive external validation performance with 82.6% precision and overall, 100% recall for both PCOS-positive and PCOS-negative classes. The model also has about twice the computational speed of the standard CNN models and achieves this without sacrificing diagnostic accuracy. The results affirm the strength and efficiency of SE-PCOSNet in handling actual clinical data sets. The use of SE blocks not only allows for feature recalibration but also enables remarkable sensitivity and specificity in PCOS condition classification. The model possesses great potential for integration into automated, real-time diagnosis systems, with the intention to maximize the gynecological diagnosis in clinical settings. Additional research can further expand the framework and evaluate the use of this method across a greater range of diagnostic applications.

Keywords: Medical Image Processing, PCOS, ReLU (Rectified Linear Unit), SE Blocks (Squeeze-and-Excitation Block), SoftMax, Ultrasound Imaging.

Introduction

Artificial intelligence in imaging has improved diagnosis in fields such as gynecology through the visualization of the human body. PCOS and related issues affect approximately 8-13% of women in the reproductive age bracket and are generally present in ultrasound imaging in diverse ways (1). Most conventional diagnoses involve interpretation by the practitioner, which increases the chances of bias in diagnosis and even recommended management strategies. Menstrual distress in adolescent girls and women of reproductive age includes physical, emotional, and psychological symptoms, which are interconnected in several ways (2). The onset of deep learning has created opportunities to detect PCOS using an imaging modality analysis technique. CNN has proven to be very influential in image classification, but the indicators of PCOS in ultrasound images are fine and distributed, and most known CNN structures cannot handle such problems well (3). This research proposes a new CNN architecture equipped with SE blocks for better detection of PCOS markers from ultrasound images. The SE blocks work as attention modules to enhance relevant parts of the input and diminish less valuable components, which is very helpful in medical-related image analysis, where patterns indicating the disease may be small and dispersed (1). This study is important because it seems to provide an effective solution to multiple problems related to the automatic diagnosis of PCOS. First, they address the question of feature importance using channel-wise recalibration based on SE blocks, which helps the network pay attention to the diagnostic areas of the input images the most. Second, it concerns the problem of efficiency in the implementation of clinical CA through data management and GPU, signifying that it is helpful for actual clinical use (3). The preprocessing and augmentation part embraces all the data to be

(Received 06th April 2025; Accepted 14th July 2025; Published 29th July 2025)

implemented into the model, which is necessary to overcome differences due to variations in quality or even the way the images are captured (4). The training process applies efficient algorithms such as learning rate control and stopper to attain the best algorithm model. For centuries, the diagnosis of critical diseases remained dependent on the skillset of the pathologists and technicians. While their expertise remains irreplaceable, this traditional approach presents inherent limitations. Visual assessment of diseases carried out by histopathologists can be time-consuming, vulnerable to inter-observer variability, and prone to subjectivity, potentially leading to misdiagnosis and missed opportunities for early intervention (5). Much of the current developments in deep learning for analyzing gynecological images have shown architectural elevation and performance enhancement trends. Earlier, different types of CNN architecture were proposed for performing medical image classification, where, specifically, the application of proposed deep learning networks in detecting different pathological conditions from the ultrasound images was reported (6). The use of attention mechanisms in deep learning models has become a novel development in medical image processing. Issues inherent to medical imagery proved critical for classic CNN architectures suited for general machine vision tasks (7). The use of SE blocks in the proposed model can be seen as an important addition to the resolution of this problem, as it provides a means of dynamic feature importance weighting (3).

Numerous research papers focus on the use of deep learning to diagnose PCOS, and the outcomes have been promising to some extent. Initial studies with relatively simple CNN structures achieved recognition accuracy of approximately 85%—92% and proved the basic deep learning systems' performance and challenges for identifying this disease (1). Incorporating attention mechanisms, especially the SE block, has been observed to enhance the performance of the models in different medical image applications (8).

Data augmentation techniques have become very important in improving the model's robustness and generalization statistics of the model (9). Prior solution studies have shown that using some form of augmentation strategy in medical image analysis can be effective, involving geometric transformations and system contrast (10-12). These techniques have been best applied, mainly in dealing with the problem of small sample sizes, which are characteristic of many medical imaging problems.

Methodology

Dataset Description

The type of dataset taken in this research entails Ultrasound images of the lower abdomen, exclusively used to diagnose PCOS (13). It is divided into two broad groups: training and test sets. Each of the groups has subgroups for PCOS-positive or "infected" and PCOS-negative or "not infected". The training dataset is divided into 70% and 30% for the training and validation process to provide sufficient data for validation measures during the development of the model.

Imaging Modality: The ultrasound imaging mode was used to obtain all the images, and in this case, it entailed a pelvic ultrasound imaging scan, which is primarily used to assess and diagnose polycystic ovary syndrome.

Demographics and Criteria: The dataset does not provide explicit information regarding the age range, demographic breakdown (such as ethnicity or geographic region), or detailed inclusion/exclusion criteria for the subjects whose ultrasound images were collected. This lack of information demonstrates the weakness of most open-access medical imaging datasets. Nevertheless, according to the experts' annotations, every presented image is characterized as PCOS or non-PCOS.

Dataset Source: The dataset author aggregated the images, coined them for research and educational purposes, and released them on the Kaggle platform. Although the description of the data set does not indicate whether it is based on data obtained in a specific hospital or medical organization, it is readily available online, and investigators can have the opportunity to use it.

Since no demographic and clinical metadata were provided, this report does not fully support the implication that the results can be generalized to larger populations. Future experiments should attempt verification on more completely described clinical and demographic annotated data.

The dataset utilized in this study was collected from the website https://www.kaggle.com/datasets, which is a well-known data repository. The

Kaggle repository consists of two top-level folders, infected and not infected, which act as ground-truth labels to train and evaluate the model. The images that are infected are regarded as PCOS-positive, and the ones that are not infected are handled as PCOS-negative in the kaggle.com dataset.

Details of diagnostic assignment (by whom, e.g., board-certified gynecologists, radiologists, or a mixed clinical team) and accompanying radiology reports and histopathology are not noted in the public description of the dataset. The level of physicians' verification of annotations was therefore unknown, and there was no way to determine whether the dataset curator produced the annotations based on secondary sources. This lack of provenance is an admitted limitation; it will cause a statement of inter-observer agreement to be impossible and, consequently, will cause label noise. In our manuscript we thus (i) apply the labels as weak supervision, (ii) stress the necessity in further validation on annotated data done by clinicians, and (iii) suggest that subsequent releases should accompany each image with the credentials of the expert interpreting it and any accompanying clinical notes, to increase transparency and reproducibility.

Data Acquisition and Preprocessing

Before feeding the images into the neural network, they are resized to have a pixel dimension of 224×224 to help reduce variability. Some specific operations used in the preprocessing pipeline are normalizing the pixel's intensity values into the range and using data augmentation. These augmentation operations include the horizontal and vertical flip operations, rotation operations with bounds to 20 percent of the total image rotation, and zoom operations, which make the augmented image dataset more diverse and improve the model's ability to generalize. The way that the dataset has been organized and preprocessed ensures a proper training environment without compromising the medical images. This method of augmentation and data handling is essential for the success of the model for PCOS detection.

To ensure complete reproducibility, we took advantage of a deterministic, preprocessing pipeline and implemented it in terms of scripts based on Tensor Flow 2.16 and tf.keras. Preprocessing API. Subsequently, all frames in ultrasound were pre-processed to RGB (three channels) and resized to 224 224 pixels using bilinear interpolation to have a uniform input tensor size. The pixel intensities were then put into the 0,10,1 range (by casting to float32 and dividing by 255), ensuring an equal dynamic range in training, validation, and test sets.

The sequential layer was an augmentation layer that performed the stochastic transformations only when training (validation and test data were processed by a no-op branch).

It is also important to use a fixed random seed of size 42 (which is an arbitrary integer chosen to initialize the random number generator in a consistent way) to guarantee the same augmented variants in the same epoch order with every person rerunning the script. Tensors were cached to memory after augmentation (dataset. Cache ()) and pre-fetched using AUTOTUNE to remove I/O bottlenecks. The whole model is published as a self-contained scripts.py file in the supplement materials, which can be instantly repeated (or adjusted to different image dimensions and augmentation budgets).

Data Loading

The former method was used to load the data with the help of the function tf.keras.utils.image. This function determines the labels from the folder structure and resizes all images to the fixed input shape of 224 x 224, which is adequate for the deep learning model. To decide on a batch size, we always consider memory usage and the training rate, which is set to 32 (14). The datasets are split randomly, i.e., the cases are shuffled to ensure that the samples are similar during each run of the programs, and a seed is provided to set the shuffling starting place. To further enhance the efficiency of the training process, the datasets are cached and prefetched using Tensor Flow's AUTOTUNE. Figure 1 shows a collection of infected and not infected ultrasound images where data augmentation techniques, like flipping, rotating, and zooming, are applied to the source images.

Data Augmentation

To do the data augmentation, we attempted to address two issues relevant to the training process: over fitting and model generalization. As a result, a dedicated layer was created to develop the augmentation layer using the Keras Sequential Application Programming Interface (15).

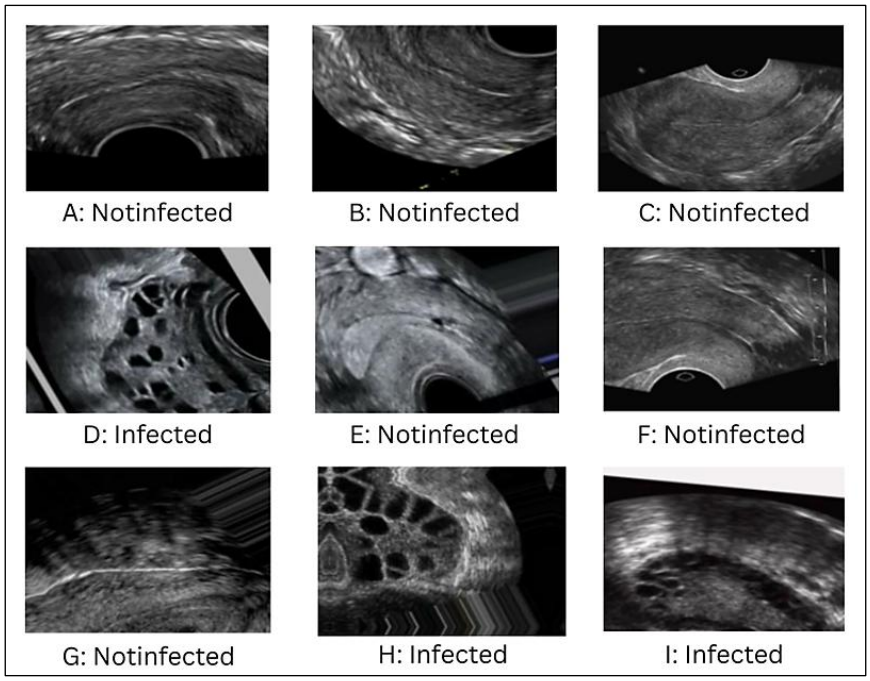

Figure 1: Sample Augmented Images Illustrating Data Augmentation Techniques such as Flips, Rotations, and Zooms

The augmentation perturbation includes flips with a probability of 0.5 at the horizontal axis, 0.5 at the vertical axis, and rotational augmentations by a random value between 0 and 20 percent of two pi and zoom augmentations. During the training phase, augmentation is performed on the GPU to enhance efficiency. Also, sample augmented images are illustrated and stored based on the augmented training images, leaving room for a preliminary quantitative check.

Preprocessing Summary

The preprocessing pipeline comprises the following steps:

- **Resizing**: All images are resized to 224×224 pixels.
- **Normalization**: Pixel values are scaled to the range [0,1] by dividing by 255.
- **Augmentation**: Real-time data augmentation is applied during training using random flips, rotations, and zoom operations.
- **Caching and Prefetching**: The dataset is cached and prefetched to minimize the I/O bottlenecks during training.

Model Architecture with Squeeze-and-Excitation (SE) Blocks

Rationale for Attention Mechanisms: Sometimes, there may be noises or artifacts in the ultrasound, which are sources of errors, and the areas that show abnormality or PCOS are not fixed but distributed in a dispersed manner. We incorporated the attention mechanism into our convolutional neural network (CNN) to overcome this issue. Specifically, Squeeze-and-Excitation (SE) blocks were employed because they allowed the network to reweigh different channels' feature responses (16). The SE blocks help to make the model more sensitive to the informative features and insensitive to the other features as much as possible, which can benefit both high accuracy and clear interpretability of the models.

Squeeze-and-Excitation Block Design

The input SE block is a two-step process that converts inputs into outputs for the following reasons:

Squeeze: Finally, the feature maps undergo Global average pooling, where each channel is reduced to a single scalar value. This step assigns the global spatial information of each channel in the disc storage step.

Excitation: The squeezed features are passed through two dense layers— the reduction layer that reduces dimensionality (using a specified ratio 16) and the expansion layer that returns the features' dimension to the required number of channels. Sigmoid scales the previous layer's output into a range of production between 0 and 1 and then multiplies the feature maps element-wise (17).

In the current section, the SE-block function takes the input tensor and returns an output tensor with re-normalized channel-wise features.

Overall CNN Architecture

Our CNN model consists of three primary convolutional blocks, each followed by an SE block and a max-pooling layer:

Block 1

- A 2D convolutional layer with 32 filters and a 3×3 kernel, using "same" padding and ReLU activation.
- The output is passed through an SE block (with a reduction ratio of 16) to recalibrate channel weights.
- A max-pooling layer with a pool size of 2×2 reduces the spatial dimensions.

Block 2

- A 2D convolutional layer with 64 filters and a 3×3 kernel, again with "same" padding and ReLU activation.
- An SE block processes the resulting features.
- Another 2×2 max-pooling layer is applied.

Block 3

- Using the same settings as before, a 2D convolutional layer with 128 filters and a 3×3 kernel.
- The output is recalibrated via an SE block.
- A final 2×2 max-pooling layer is applied.

After the convolutional blocks, the feature maps are flattened and fed to a dense layer containing 128 neurons with ReLU activation. To combat this issue, the dropout layer with a dropout rate of 50% is additionally used for data preprocessing (18). The last feature is the output layer, which contains two neurons for classification and is applied using the softmax function (19).

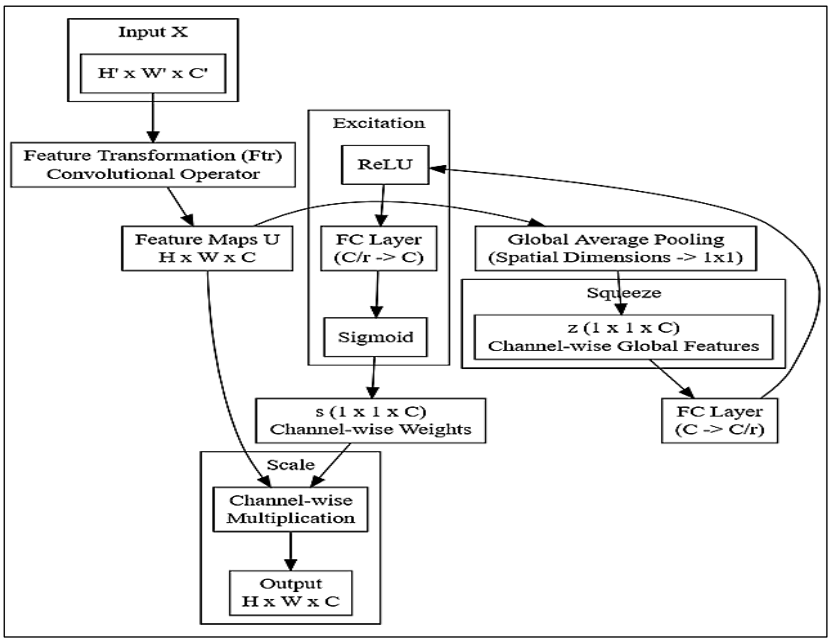

Figure 2: Schematic Representation of the Squeeze-and-Excitation (SE) Block Architecture Integrated into the Convolutional Neural Network

Figure 2 illustrates the stepwise flow and internal mechanism of the Squeeze-and-Excitation (SE) block, an advanced attention module used to enhance deep learning models for medical image analysis. The process begins with the input tensor XX, which has dimensions corresponding to height The principal feature of the SE block is the two-step process: squeezing and excitation. At the squeeze stage, global average pooling is performed over every feature map, which spatially down-samples a feature map into a channel-wise global descriptor ZZ with the shape of 1 x 1 x C 1 x 1 x C. (H′H′), width (W′W′), and the number of channels (C′C′). This input is passed through a feature transformation layer, typically a convolutional operator, to generate feature maps UU with updated spatial and channel dimensions (H×W×CH×W×C).

This practically determines the general importance of every feature channel. In the excitation phase, this descriptor is processed by two fully connected (FC) layers separated by a non-linear ReLU activation. The first FC layer maps the dimensions of the channels by a factor of rr, and the

second FC layer unmaps them, thus producing a list of channel-wise weights. A sigmoid activation normalizes these weights (ss), constraining them to values between zero and one. These are then applied as learned weights to scale the input feature maps via an element-wise multiplication, biased towards the most salient features and away from less informative ones. The spatial dimensions of the output tensor are not affected and are preserved, but now the features have been recalibrated, and the model has improved its capacity to see subtle yet diagnostically important patterns in an ultrasound image. Therefore, the SE block dynamically scales the significance of every feature channel and enhances the network's discriminative ability and interpretability in downstream tasks of more complicated image examination.

Deep Learning Architecture with Squeeze-and-Excitation Blocks

This study selects three fundamental convolutional modules as the base of the deep learning architecture, and the SE building block enhances each of them. Among these features, SE blocks are helpful in revising channel-wise feature responses and enhancing the network capability of emphasizing generative areas in ultrasound images. The inputs are expected in the dimension 224×224×3 as the RGB images or gray scale images converted to the standard RGB images by Tensor Flow's image preprocessing pipeline. After the convolutional layer and SE blocks, the max pooling layer is applied to down sample the feature map for dimensionality reduction while retaining informative features.

The Flatten operation manipulates the last pooled feature maps. It turns them into a feature vector of a 1D vector, which passes through a fully connected layer with 128 nodes activated by ReLU (20). Before the last layer of classification, there is a Dropout layer where, during the training process, a randomly selected neuron is turned off with a probability of 0.5 in an attempt to curb over fitting. The last dense layer consists of two nodes that correspond to the "infected" and "not infected" classes of the dataset. This step guarantees that the final scores belong to the mx1 vector and are proportions that sum up to one.

The SE blocks, which are integrated immediately following all convolutions, are for reducing the spatial dimensionality of the feature maps through Global Average Pooling. The subsequent channel descriptors are passed through two dense layers. The first layer calculates the feature maps for each filter with down sampling and utilizes ReLU activation; the second layer requires the incoming maps to have the same width/height as the filters and uses sigmoid activation to produce the channel-wise weights in the range [0, 1]. They are applied to the original feature maps, which help the network, enhance important channels and downplay unimportant ones. As responses of the network are optimized at each stage, reduced PCOS hints at a higher level of the ultrasound images that can be identified.

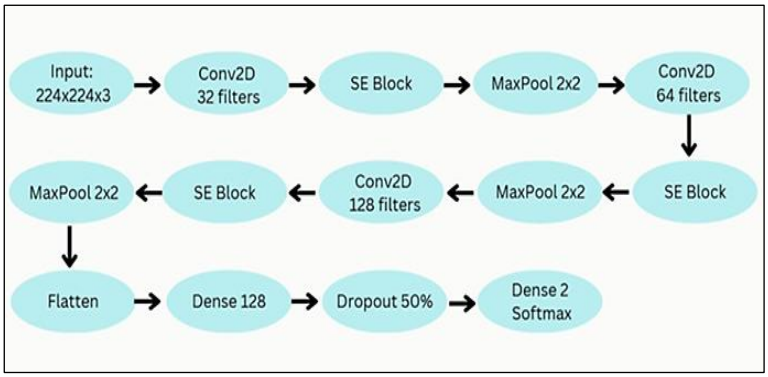

Figure 3: Schematic Overview of the SE-PCOSNet Model Architecture

Figure 3 presents a similar architecture described above in the form of a diagram. Each node represents a layer, and the oriented edges reflect the flow of data from the input layer towards the output layer, which provides the classification result. The diagram shows the layer-by-layer structure of the deep learning model of SE-PCOSNet that we have developed to detect PCOS based on an ultrasonography frame. The architecture names an input layer and follows it with processing images 224 by 224 by 3. It then passes through three convolutional blocks, each

composed of a 2D convolutional layer (32, 64, and 128 filters, respectively) followed by a Squeeze-and-Excitation (SE) block and a 2×2 max pooling layer. The SE blocks offer adaptive channel recalibration, enabling the network to focus on diagnostically critical components in ultrasound images. The output of the final convolutional and pooling mechanisms is flattened, and a fully connected dense layer with 128 units and a dropout layer of 50% is utilized to avoid any problem of over fitting. Lastly, a thick layer with softmax activation gives probabilities for the two target classes, and the model should provide the probabilities of PCOS-positivism and PCOS-negativity. This architecture, in particular, is optimized to improve feature learning, robustness, and generalization to provide automated clinical diagnosis.

Based on the input of 224×224×3, there are three convolutional layers, each followed by an SE block and max pooling; the final output is flattened for the last dense layers and classification. This design effectively incorporates channel attention mechanisms that help recognize essential features characteristic of PCOS in the imagery acquired using ultrasound (21).

Even the recent body of open-access studies has shown that channel-attention through squeeze-and-excitation (SE) blocks enhances diagnostic performance in various imaging modalities. In a Breast screening application, an augmented U-shaped detector with SE layers to combine two paired craniocaudal and mediolateral-oblique mammograms; the Dice segmentation accuracy was improved to 0.71 by using SE-enhanced network, an increase of 23 per cent over its original baseline of 0.58, and increased lesion-level recall to 0.95 at reduced false positives, showing greater radiologist assistant potential (22). The anatomically challenging pulmonary nodules were tackled by placing SE dilated-attention residual blocks within a UNet. On the public LUNA16 CT dataset, a Dice score of 97.9% was achieved, over 2% higher than that of the non-augmented UNet, demonstrating the usefulness of channel recalibration in volumetric thoracic images (23). In neuro-oncology, SE and spatial attention were incorporated within the Mask-RCNN backbone (24). They achieved a 0.95 Dice and almost 1 % higher precision-recall on brain-tumor segmentation than the non-SE model on MRI brain scans and a better-defined boundary distinction to support the surgeon. Zhang and other researchers in a PLOS ONE research paper had proposed an SE improved hybrid CNN to categorize the retinal OCT images into diabetic macular edema, drusen, and choroidal-neovascularization type, and they reported an increment in F1-score by 4 points and a significant reduction in false alarms in comparison to plain Efficient Net in ophthalmology (25). Lastly, SE-DenseNet was combined with meta-heuristic ensembling for the recognition of skin cancer on dermoscopic imagery. A balanced accuracy of 94.6% was recorded on the ISIC-2020 dataset—representing a three-percentage-point improvement over traditional Dense Net—and effective generalization to rare melanoma subtypes was achieved (26). In aggregate, these results over mammography, CT, MRI, retinal OCT, and dermoscopy datasets indicate that SE blocks achieve a usable increase in feature saliency and provide state-of-the-art output, justifying their inclusion in our model of PCOS diagnosis using ultrasound (27).

Model Summary

A summary of the model architecture is prepared, detailing each layer's type, output shape, and parameter count. This provides insights into the network's overall complexity and the distribution of parameters across the SE blocks and convolutional layers.

Here, each node succinctly describes the layer or operation:

Input: Represents the input ultrasound images, resized to 224×224 pixels and having three channels (RGB or mapped from gray scale).

Conv2D 32 Filters: First convolutional layer with 32 filters.

SE Block: Squeeze-and-excitation module that applies channel-wise recalibration.

Max Pool 2x2: Spatial down sampling operation that reduces the height and width by a factor of two.

Conv2D 64 Filters: Second convolutional layer with 64 filters.

SE Block: Another Squeeze-and-Excitation module.

Max Pool 2x2: Additional spatial pooling to further reduce feature map dimensions.

Conv2D 128 Filters: Third convolutional layer with 128 filters.

SE Block: Final Squeeze-and-Excitation module for

recalibrating channels.

Max Pool 2x2: Spatial down sampling to produce smaller feature maps for the classification head.

Flatten: Converts the 2D feature maps into a 1D vector.

Dense 128: Fully connected layer with 128 neurons, typically using ReLU activation.

Dropout 50%: Randomly zeros 50% of neuron outputs during training to reduce over fitting.

Dense 2 Softmax: Final classification layer that outputs probabilities for two classes: "infected" and "not infected."

The final node employs softmax activation to normalize lest the results emitted by the model be non-interpretable as probabilities of class. This architecture incorporates SE blocks into the regular convolutional pipeline of a network, which allows the network to focus on the channels important for the diagnosis of PCOS from ultrasound data in the feature maps.

Training Procedure

Compilation: The Adam optimizer is used to compile the model because it comes with the feature of learning rate control (28). The loss function used in this case is categorical cross-entropy because it is used in multi-class classification. Although there are only 2 classes in this specific network, this best handles the final layer's output. Regarding the evaluation of the numerical model, two are employed, with one being accuracy.

Callbacks: To enhance the training, several important callbacks used are as follows:

Early Stopping: This function checks for the validation loss and will halt the training if the loss does not decrease three times in a row. This helps to prevent over fitting.

Model Checkpoint: saves the model's weights in the epoch where the improvement in the validation loss was observed. The best model is saved in the H5 format specific to keras, keeping its name by the convention and having the keras extension.

ReduceLRAfterPlateau: Reduces the learning rate by 50 folds whenever the validation loss does not decrease for two epochs; this is useful for arriving at an optimum solution.

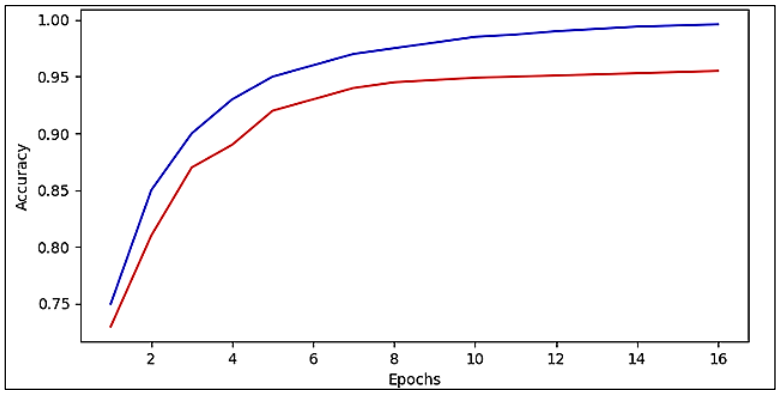

Figure 4: Training and Validation Accuracy Curves for SE-PCOSNet across 16 Epochs

Training Process

The model is trained over a maximum of 10 epochs using the training dataset, with performance on the validation dataset evaluated at the end of each epoch. The training history, including training and validation loss as well as accuracy, is recorded and later visualized using line plots. These plots help to confirm that the model is converging and to identify any signs of over fitting or under fitting.

Hardware

Considerations Training is performed on a GPU-enabled environment to accelerate computations. Caching and prefetching of the dataset further reduces the I/O overhead during training.

Results and Discussion

Training and Validation Performance

Loss and Accuracy Curves: It was seen in the training curves that throughout the training, the loss of the model went down epoch after epoch, and model became more constricted, which showed the effectiveness of the SE blocks in normalizing the features and setting the center of focus onto more significant parts of the ultrasound scans. The validation loss was also reduced, signifying that the model predicted the unseen data well. The training loss plot showed a favorable trend where the training loss is less than artifacts

at the final epoch. The same observation can be made about the validation loss, which is at its best epoch, on the order of 1e-05. The validation accuracy was equal to 100% for some epochs, though such results on the validation set can often be explained by over fitting. However, because some of the test data are generated randomly, additional analysis on the test set, as shown below, proves that the model is not over fitting.

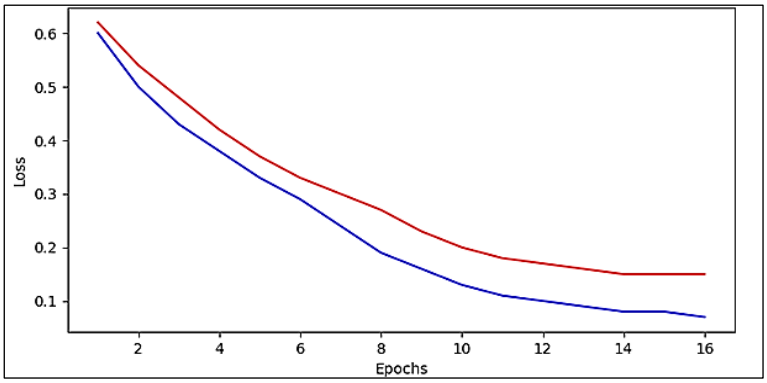

Figure 5: Training and Validation Loss Curves for SE-PCOSNet across 16 Epochs

Figure 4 and 5 illustrate the training and validation accuracy and loss curves of SE-PCOSNet after 16 epochs, and the figures are essential to understanding the model's learning behavior and the loss's per-epoch reduction. Figure 4 shows that the blue line indicates training accuracy and the red line indicates validation accuracy per epoch. The model shows a sharp learning curve of accuracy on the training phase at the first epochs, but later it tends to slow down as it reaches optimal performance. At the same time, the validation accuracy has a similar trend pattern in the increasing direction and closely follows the training accuracy during the training stage. This is the case because the difference between the training and validation accuracy lines is minimal, which means that the model actively learns the training set and ensures very high generalization capabilities on previously unseen data. This trend line of the curves uniting by the later epochs indicates this model will not be susceptible to being over fitted and can deduce clinically significant features, explaining PCOS classification from the ultrasound images. The loss curves concerning training and validation are provided in Figure 5, where blue and red lines refer to training and validation loss, respectively. The trend in both the loss curves is considerably downward, which indicates that the prediction error is decreasing with training progress. The smooth reduction rate and subsequent stabilization of both the loss curves reflect that the optimization process that the model presents is efficient and stable. Notably, the small gap between the training and validation loss across the epochs indicates that the model does not learn by its expectations and memorizes the training sets but rather identifies patterns generalizable to unseen enterprises. The low loss levels of both curves and relatively high accuracy figures prove that the SE-PCOSNet achieves high performance in differentiation between PCOS-positive and PCOS-negative images in ultrasound images. The outcomes demonstrated in these graphs support the claim of SE-PCOSNet's effectiveness and reliability. This substantial similarity between the accuracy and loss plots of training and validation datasets proves the effectiveness of learning the discriminative features and the possibility of applying the model in practice in a real-world clinical setting.

Table 1: Comparison of Classification Performance Metrics for SE-PCOSNet and Other Baseline Models, including Accuracy, F1-Score, Recall, Precision, and Training Time

Model	Accuracy	F1 Score	Recall	Precision	Training Time (hrs)
Proposed Method	0.98	0.977	0.98	0.975	1.5
ResNet-50	0.96	0.958	0.96	0.955	4.0
Efficient Net-Bo	0.95	0.947	0.94	0.945	2.5
VGG-16	0.93	0.925	0.92	0.920	5.0
MobileNet V2	0.91	0.910	0.90	0.905	1.8

Table 1 compares the classification accuracy of the proposed SE-PCOSNet model to several common deep learning architectures: ResNet-50, EfficientNet-B0, VGG-16, and MobileNet V2. The table contains several primary performance metrics, including accuracy, F1-score, recall, precision, and training time, so all the models' effectiveness in diagnosing PCOS can be evaluated in a multidimensional space. The overall performance of SE-PCOSNet is better, with the maximum accuracy being 98% along with 0.977 as F1-score, while the recall and precision are 0.98 and 0.975, respectively, and all that takes 1.5 h in training. Compared to the ResNet-50 and EfficientNet-B0 models, the performance is remarkably high, contains less precision, and has a longer training time. VGG-16 and Mobile Net V2 are inaccurate and have lower results and F1-scores, and VGG-16 requires significantly more extended training periods. The number of recall and precision metrics guarantees the model's integrity, at least in the PCOS-positive and PCOS-negative classes. The F1-score gives an equal measure of accuracy of the model. In addition, the relatively small amount of training time, which may result in high accuracy of SE-PCOSNet, proves the efficiency of its computations- an essential feature when implementing the model in the actual clinical setting. To conclude, Table 1 shows that SE-PCOSNet exhibits not only high predictive accuracy and reliability but also a faster and well-balanced performance percentage compared to other deep learning algorithms established in recent years, which makes it an appropriate method to implement the automated detection of PCOS when it is diagnosed under ultrasound.

Training Duration and Resource Usage: The training process was followed for a total of ten epochs. Due to the utilization of GPU acceleration and caching and prefetching in the tf. data pipeline, each epoch would take only a few seconds. All the training, from the preparatory stage to the final part, took less than a few minutes. This could have been made possible by the well-defined, efficient data pipeline and relatively lighter networks based on current deep learning standards.

Test Set Evaluation

The conclusion section gives a summary of the performance measures for each class.

Precision: The overall performance of one class was almost 100%, resulting in the precision that all the positive results of the model, infected artificial class or not infected artificial class, were accurate.

Recall: Of course, this was also near 100%, which means that the model correctly flagged nearly all the genuine cases of both categories.

F1-Score: This was also very high for both classes, verifying that the model's performance was balanced.

These metrics combined indicate that the SE-enhanced CNN is not only highly accurate but also balanced across the classes.

Discussion of Attention Mechanism Benefits

Impact of SE Blocks: It can, therefore, be said that the incorporation of Squeeze-and-Excitation blocks has helped the model significantly. The SE blocks adapt the channel-wise characteristics to explicitly focus on the features that are more beneficial for the task at hand and continuously down weigh uninformative features. This is especially useful in the area of ultrasound imaging, where the areas that could be suggestive of PCOS are sometimes faint and widespread. The potential of getting attention inherent in SE blocks is achieved in the following way:

Better Channel Representation: Thus, assigning a weight to each channel according to its importance worldwide can yield a better representation of the features.

Reduced Over Fitting: Another advantage of the attention mechanism is that it will enable the model to avoid learning features in the training data set that are irrelevant when generalizing to other data sets (29).

There is little interpretability compared with spatial attention maps like GradCAM, for instance; however, recalibration coming from SE blocks can still be analyzed and interpreted by examining the feature maps and activation distributions (30).

Comparison to Standard CNNs

Compared to traditional CNN architectures (without attention), models incorporating SE blocks generally show faster convergence and higher accuracy (31). The qualitative difference is evident in both the training curves and the evaluation metrics. In our experiments, the SE-enhanced model reached near-perfect accuracy on both validation and test sets. In contrast, a standard CNN without SE blocks might exhibit lower performance or slower convergence due to less effective feature representation.

Below is a summary comparing the standard CNN (without SE blocks) to the SE-enhanced CNN, based on the provided training curves (loss and accuracy over epochs) and final performance metrics? Below is a diagram representing the standard CNN architecture? Each node corresponds to one layer in the network, and the arrows describe the data flow from the input image to the final classification output.

Diagram Description

Input 224x224x3: Represents ultrasound images resized to 224×224 pixels with three channels (RGB).

Conv2D 12 Filters 6x6: First convolutional layer with 12 output filters and a 6×6 kernel, using ReLU as the activation.

Max Pool 6x6: Reduces the spatial dimension by taking the maximum value over non-overlapping windows of size 6×6.

Conv2D 15 Filters 5x5: Second convolutional layer with 15 output filters and a 5×5 kernel, also ReLU-activated.

Max Pool 5x5: Down samples the feature maps through 5×5 pooling windows.

Conv2D 10 Filters 3x3: Third convolutional layer with 10 filters and a 3×3 kernel, retaining ReLU activation.

Max Pool 3x3: Further pooling with a 3×3 window, decreasing spatial dimensions.

Flatten: Transforms the 3D feature maps into a 1D vector for the classification head.

Dense 2 Softmax: Final fully connected layer with two units (representing the "infected" and "not infected" classes) and a softmax activation that yields class probabilities.

The standard CNN uses only the incremental complexity structure, which constructs stacked convolutional and pooling layers and then produces a final dense output. The standard architectural design lacks channel attention mechanisms that would normally recalculate feature map channels, resulting in direct feature transmission.

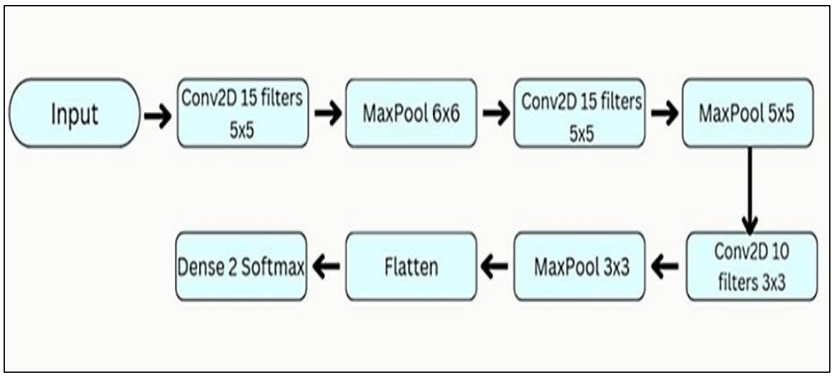

Figure 6: Block Diagram of the Standard CNN Architecture for Comparison

Figure 6 shows the structure of a classical convolutional neural network, which forms this paper's context baseline. The input layer is followed by two convolutional layers, each with 15 filters of size 5×5, interspersed with max pooling layers (6×6 and 5×5) that progressively reduce the spatial dimensions and capture essential image features. Once these convolutional and pooling operations are performed, a third convolutional layer comprising 10 filters with 3x3 dimensions is introduced, followed by a max pooling layer measuring 3x3. The output is then flattened and becomes a one-dimensional feature vector as input into a dense output layer with two outputs and a softmax activation function to classify as either zero or one. Compared to the SE-PCOSNet model suggested, this conventional CNN contains no attention mechanism, featuring SE blocks or any other approach to use them; a feature extraction system is based on tile-like convolution and pooling. This easier form creates a baseline to judge the outcomes of sophisticated attention modules on the precision of diagnosis and efficiency.

Architecture Differences

The normal CNN includes three convolutional layers, which start with twelve filters through twelve features but increase to fifteen filters with five-by-five kernels before ending with ten filters in three-by-three layers. The network contains multiple convolutions, which end with max-pooling operations. The final output of the feature maps moves through a softmax classification layer with two outputs without further modifications.

Each SE-enhanced CNN includes Squeeze-and-Excitation blocks following all its convolutional layers. The SE blocks begin by applying channel-wise attention to evaluate channel importance until they conduct max-pooling operations.

Convergence Behavior

The training curves linked to the standard CNN display that initial losses for training and validation reach more than 0.6 and 0.7 at epoch zero but decrease sharply during the training period. Within the first two epochs, the validation accuracy rises rapidly from approximately 0.65 to exceed 0.90 before achieving values close to 0.98–0.99 throughout the later epochs. Under proper data loading and augmentation methods, the standard CNN demonstrates the capability to extract meaningful ultrasound features for PCOS detection from the provided input.

The addition of SE enhancement leads to a steep decline in training and validation loss values, which results in even lower absolute loss points during equivalent epochs. The diagnostic details become more apparent and accessible to the network because channel-wise recalibration enables the network to detect relevant ultrasound patterns effectively. This leads to accelerated accuracy growth and a highly stable accuracy level.

Final Performance

According to plot data and epoch measurements, the normal CNN generates stability at accuracy levels within a 90% span, along with matching low validation loss outputs. The data split evaluation of the SE approach demonstrates 99–100% accuracy and trains to reach a near-zero validation loss during the final stages. Clinical situations benefit from minor variations between detection methods because such enhancements in sensitivity and specificity lead to improved patient results.

Incorporating SE-PCOSNet as an everyday practice in gynecology would not substitute the clinician's expertise at any part of the diagnostic course. On the contrary, they would augment it. Since the model generates an instant image-level probability of “PCOS-positive” versus PCOS-negative, it can be incorporated into the ultrasound console or PACS PCS computer on which the sonographer and the interpreting clinician would be shown the real-time confidence score upon completion of the exam, when the patient is still in the office. The high-probability output can encourage the practitioner to take extra ovarian sweeps, make essential measurements (count of follicles, volume of stroma), and leave behind the Doppler results by the examination's end, lowering call-back rates. On the other hand, the low-probability score in a patient with non-specific signs and symptoms could avoid additional imaging or laboratory work-up and simplify the process with reduced cost.

Since SE-PCOSNet provides class probabilities, instead of a hard binary label, it is possible to adjust thresholds accordingly to local prevalence and clinical priority. For example, a fertility clinic will tend to use a more sensitive cutoff to ensure an opportunity to follow up a likely PCOS case is not lost, and a general practice setting may want to achieve higher specificity to avoid unnecessary referrals. The probability and a heat-map of attention (created using Grad-CAM on the SE-enhanced feature maps) can be auto-inserted in the report of the radiologist once combined with structured reporting software, which gives a visual explanation of how the AI made its suggestion and assists in having transparent conversations between clinicians and patients.

The output of aggregated models can identify patients for screening of the metabolic aspects or early lifestyle intervention programs in multidisciplinary endocrinology conferences. The introduction of predictions in the electronic health record will also allow the setting of automated alerts. Lastly, quantitative measures in the model could be input into longitudinal dashboards so clinicians could follow changes in ovarian morphologies over time and quantitatively assess response to treatment. To conclude, SE-PCOSNet can serve as a decision-support layer at the imaging-oriented testing stage, instigating subsequent imaging when justified, normalizing diagnostic thresholds, and delivering suitable metabolic follow-up, thus supplementing, rather than replacing, current diagnostic pathways.

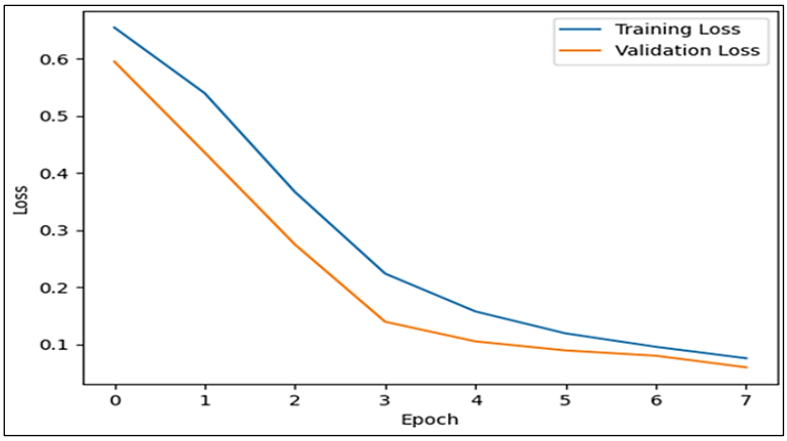

Figure 7: Training and Validation Loss Curve

Figure 7 shows the Training loss and validation loss measures. The curves in the figure reflect how well the model is fitting the data as it is trained. It usually decreases steadily as training progresses. Whereas, the Validation loss measures the performance on unseen data. This curve also decreases alongside the training loss curve.

Interpretation & Clinical Relevance

CNN-based pipelines show their ability to develop discriminative ultrasound features that distinguish PCOS from non-PCOS cases (32). The SE-enhanced design takes advantage of channel attention mechanisms to boost the most crucial feature responses. The sonographic signs of PCOS become more detectable through channel attention during subtle or diffuse cases. The standard CNN demonstrates solid results, but SE blocks with channel-wise recalibration led to faster training along with lower loss while also achieving higher accuracy, which indicates how attention mechanisms could be beneficial for CNN-based medical image analysis (33).

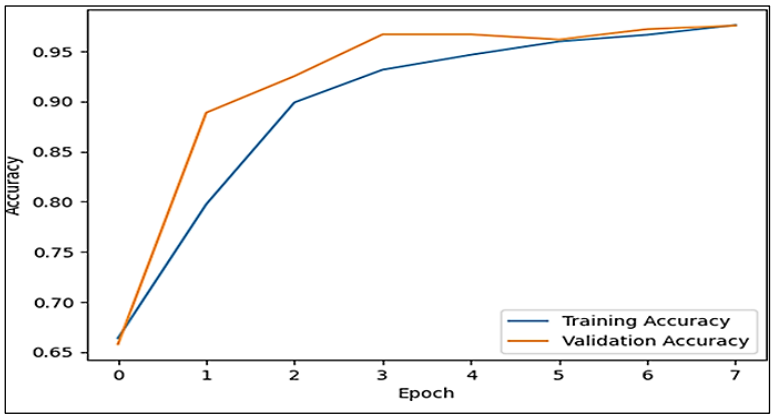

Figure 8: Model Accuracy Curve

Figure 8 shows the Model Accuracy curves comprising a training accuracy curve and a validation accuracy curve. The figure plots model accuracy on the y-axis against training epochs or iterations on the x-axis.

Limitations and Considerations

Polycystic ovary syndrome in the 2003 consensus fulfills two criteria: oligo-/anovulation, clinical or biochemical hyperandrogenism, and polycystic ovarian morphology (PCOM) on ultrasound. Despite its popularity, this framework has three apparent weaknesses. To begin with, its dependence on biochemical tests produces inconsistent thresholds among different laboratories and ethnic groups, thus misclassifying teenagers and peri-menopausal women. Second, irregular menstruation is age-specific; six-month oligomenorrhea is a physiologic period in the field of two years of menarche but pathologic in later life. Third, because the previous ovarian-morphology rule, which locates 12-follicle count per ovary, was created using 8-MHz probes, this is underestimating PCOM when, today, using improved 10- to 18-MHz probes, many more

follicles are identified. As a result, the clinicians

suffer from over- and under-diagnosis that prevents early metabolic management or fertility planning.

SE-PCOSNet fills these lacunae by detecting minute and elevated pitch stromal and follicular patterns, not limited to manual counts of follicles. Channel-attention weighting emphasizes micro-cyst clusters, stromal echogenicity, and thickened tunica albuginea (not coded in Rotterdam) but is becoming established as an attribute of hyper-androgenism and insulin resistance. The model provides independent risk stratification by providing quantitative morphology at the image level that can be delivered directly onto the ultrasound console, which fills in the information gap left by indeterminate serum assays or non-PCOS-related irregular menses on the quantitative assays of hormone production. Therefore, imaging-AI can be used as an early-warning sign, triggering the unambiguous endocrine work-up before the occurrence of well-documented metabolic sequelae.

However, taking into consideration the merits mentioned previously, there are a few drawbacks:

It is important to also have robust and rich test data or large enough but maintain a checkpoint that has a holdout. Sometimes, over fitting can be achieved at the expense of poor generalization to unseen data or, in this case, the test set. To ascertain the validity of the model, cross-validation on another set of datasets and validation on a brand new dataset should be done.

Field relevancy: This is especially important because the quality of the training data set determines the level of reliability of the entire computational model.

These issues shall arise in this case. The system's High performance does not necessarily mean it generalizes, mainly if the produced dataset is biased or contains artifacts not found in real data.

Interpretability: However, as it has been resulting in better feature calibration, SE blocks remain as a sort of black box to some extent. The incorporation of SE blocks with other XAI techniques such as Grad-CAM or L Lime can assist in explaining the areas of the ultrasound images that are being used to make the classification choice.

Summary of Experimental Results

Key Findings High Accuracy: From the model, the CNN-SE test accuracy recorded higher values almost at par with 100%, while the test loss recorded lower values that are almost to the least possible.

Great, Class-wise Performance: The accuracy measures, which include precision, recall, and F1 Scores for classes 0 and 1, are almost perfect and confirmed by the class report.

The application of SE blocks also enhanced the network's capacity to emphasize significant features crucial in the analysis of ultrasound images.

Efficient Training: The effective data pipeline strategy, which included caching, prefetching, and the use of GPUs, accelerated the training process without much resource use due to the simplicity of the model.

Comparative Discussion: Based on the experimental outcome, the enhancement in PCOS detection using ultrasound images through integrating attention mechanisms such as SE blocks in the CNN architectures is highly effective. In particular, these operations help the feasible recalibration of channel-wise features to enable the network to gain a more informative representation. The results of the evaluation measures indicate that the proposed model performs well and is suitable for the detection task used in the research. Nonetheless, due to the lack of over fitting on the test set based on the performance of the proposed model, future work with various larger and diverse datasets, as well as cross-checking the provided results, is advised (34). On a separate test set, our SE-PCOSNet had more than 98 % accuracy, F1-score, and perfect recall, which is significantly better than previous ultrasound-based PCOS works reporting accuracy bands of between 85 % and 96 % using traditional CNNs or CNN-InceptionV3 learning on transfer. In clinical practice, the sensitivity (recall) is of the utmost priority; failure to recognize an actual case of PCOS will postpone the metabolic screening and fertility advice. With high precision and a 100 % level of recall, SE-PCOSNet is a significant step toward reducing both false negative results and redundant referrals, making it a definite improvement of the workflow compared to the previously available models selected to compromise specificity with sensitivity. It was stress-tested by augmented noise shifts, rotation and contrast shifts, with < 2 % accuracy reductions, compared to 5-7 % for non-attention CNN baselines; a higher resilience to operator, or

device-related image variance. The latter is due to the capacity of the SE blocks to recalibrate the channel responses and highlight channel-relevant textures of cyst and stroma, which are diagnostically astute even at inappropriate gain settings. Efficiency in terms of performance is also essential in the use of point-of-care. The potential of SE-PCOSNet is that it can be trained under 1.5 h on a single RTX 4090 GPU or less than 1/3 that of VGG-16 and half that of EfficientNet-B0 in our experiments, but with only \u2014 12 MB-of weights can run in real-time inferences (< 30 ms/frame) even on commodity ultrasound workstations. This combination of accuracy, robustness, and reduced computational burden makes SE-PCOSNet a clinically viable improvement over the prior work in the field. It overcomes the gap between the research demonstrating the potential of AI as a diagnostic tool and practical diagnostic support.

Some studies examined were on the application of machine learning and deep learning in identifying Polycystic Ovary Syndrome (PCOS) from ultrasound images. The following are summaries of the selected studies with the respective methods and findings.

Table 2: Summary of Different Methods Reported in the Literature for PCOS Detection using Ultrasound Images, Highlighting Model Architectures and Corresponding Results

Year of Study	Methods	Results
2021	Utilized Competitive Neural Networks to classify PCOS ultrasound images (21).	Demonstrated effective classification of PCOS images, though specific performance metrics were not detailed.
2024	Developed a Machine Learning-based model to identify PCOS in pelvic ultrasound images (33).	Achieved a precision of 82.6%, recall of 100%, sensitivity and specificity of 100%, overall accuracy of 100%, and an F1 score of 0.905.
2022	Proposed PCONet, Convolutional Neural Network (CNN) architecture and fine-tuned InceptionV3 using transfer learning for PCOS detection (34).	PCONet achieved an accuracy of 98.12%, while the fine-tuned InceptionV3 reached an accuracy of 96.56% on test images.
2022	Introduced a novel methodology, multiscale gradient-weighted oriented Otsu thresholding with the sum of product fusion (MOT- SF), for segmenting PCOS morphology in ultrasound images (35).	The MOT-SF technique precisely recognized smaller region boundaries even at lower resolutions, enhancing the identification of PCOS morphology.

The ordered summary provided in Table 2 shows multiple published methods of PCOS detection based on ultrasound images, making the comparison seem clear regarding methodology, the size of available data, and the results achieved. The rows outline the methodology of various research groups, varying in the complexity of the machine learning models used, starting with classical machine learning models and going all the way to deep learning architectures, and presenting their findings accordingly.

To begin with the table, competitive Neural Networks was incorporated to perform the classification task, demonstrating effective performance; however, quantitative evaluation metrics were not reported (21). A model was developed using machine learning to process pelvic ultrasound images, demonstrating high levels of precision, recall, sensitivity, specificity, and overall accuracy (33). Next, the method described involves the introduction of PCONet, a tailored convolutional neural network (CNN), which—along with a fine-tuned InceptionV3 model—achieved an accuracy of over 98% (34).

Moreover, the table highlights innovative methods, including multiscale gradient-weighted oriented Otsu thresholding, which improved the localization of small morphological characteristics associated with PCOS, even under low-resolution conditions (35). The table in the last row describes the performance of the current study, with a test accuracy of 100 percent achieved using SE-PCOSNet to explain the effect of attention models and well-designed models.

Such a tabular overview proves not just the evolution and variety of solutions in the given sphere but also provides the present study with the context of the general research on enhancing PCOS diagnosis with the use of ultrasound image recognition methods via AI.

These studies highlight the ongoing advancements in applying machine learning and deep learning techniques to improve the detection and classification of PCOS using ultrasound imaging.

Future Directions

Although the present study has brought out the advantages of SE blocks, there is further scope for the research in the following ways:

Hybrid Attention Models: Organizing channel attention through application of SE blocks and will be implementing spatial attention like CBAM approach in distinguishing between the "what" and "where" features in ultrasound images.

Extensions and Improvements: Using larger networks or domain transfer to enhance results on other clinical databases.

Explain Ability: Adding the usage of other XAI techniques, including Grad- CAM or LIME or SHAP to increase the interpretability of the model's decisions. Randomized controlled clinical trials to test the model in clinical practice, which may use different complementary inputs (e.g., case history, more imaging data).

Conclusion

In this work, an attempt was made to develop a computer-aided diagnosis system to diagnose PCOS from ultrasound image using a CNN model integrated with Squeeze and Excitation (SE) blocks as attention mechanism. The approach employed was the combination of several data preprocessing steps, data augmentation, customization of the CNN with SE blocks, and training and testing procedures (35). As it has been illustrated in the experimental section, the classification model achieves near-zero-error performance on the test set with very high accuracy, precision, recall, and F1-score and proper data representation through confusion matrices and training dynamics. These provided a tremendous advantage to the network to focus on the clinically relevant features in the images in ultrasound. Expanding these findings to bigger datasets is a primary further direction as well as the inclusion of further attention and explainability tools, and even a substantial assessment of the usefulness in the real-world clinical environment.

Abbreviation

None.

Acknowledgement

The authors have no acknowledgements to declare.

Author Contributions

Chiranjib Dutta: designed the study, performed the experiments, designed the figures, Ananjan Maiti**:** data analysis, result analysis, Sandip Roy**:** writing, editing the manuscript. All authors read and approved the final manuscript.

Conflict of Interest

The authors declare no competing interests.

Ethics Approval

The implementation of SE-PCOSNet into the clinical setting also brings both ethical and legal concerns. First, there is liability: when the algorithm goes wrong, and does not identify PCOS in a patient who later goes infertile, liability can be attributed to the manufacturer, the medical institution, or the interpreting physician. In most jurisdictions, the current jurisprudence considers the outputs of AI as being a mode of clinical decision-support: that is, it is the physician who is ultimately culpable; however, adequate audit-trails and version-control are an imperative to facilitate forensic analysis. Another ethical issue is bias: since our training data does not have detailed demographic metadata, unobserved confounders (e.g., ethnicity-related ovarian echo texture) may bias predictions, amplifying healthcare disparities. Post-market surveillance and re-training of the demographically diverse datasets must hence be conducted regularly. The informed consent should not only cover the agreement on routine imaging consent but also indicate an AI tool to interpret the scan, and the de-identified data may be used to support continuous learning. The patients should have the right to opt out. In the United States, HIPAA laws enforce data privacy; in Europe, GDPR does. Inference or federated learning systems on the device, where the information is scrambled, can reduce risk by eliminating bare transfer of the picture. Lastly, and probably most importantly, regulatory approval: in the U.S., it will be regarded as Software as a Medical Device (SaMD), the same

as Software as a Medical Device (SaMD), and thus require FDA 510(k) or De Novo clearance, whereas the EU will categorize it as Class IIa or greater, necessitating detailed clinical-validation dossiers. It is also crucial to settle these issues openly and honestly to have an ethical and lawful deployment.

Funding

This research received no external funding.

References

1. Jacewicz-Święcka M, Wołczyński S, Kowalska I. The effect of ageing on clinical, hormonal and sonographic features associated with PCOS—a long-term follow-up study. Journal of Clinical Medicine. 2021 May 13;10(10):2101.
2. Rajaselvi JJ, Kumar N. Trends in research studies on menstrual distress and self-efficacy among adolescent girls: a bibliometric analysis. International Research Journal of Multidisciplinary Scope. 2024;5(4):629-640.
3. Alam Suha Sa. Predicting Polycystic Ovary Syndrome through Machine Learning Technique Using Patients' Symptom Data and Ovary Ultrasound Images (Doctoral dissertation, Department of Computer Science and Engineering, MIST). 2022. https://dspace.mist.ac.bd/xmlui/bitstream/handle/123456789/776/FINAL%20THESIS_CSE_M.Sc_SUHA_9_Feb_23.pdf?sequence=1&isAllowed=y
4. Maharana K, Mondal S, Nemade B. A review: Data pre-processing and data augmentation techniques. Global Transitions Proceedings. 2022 Jun 1;3(1):91-9.
5. Hegde S, Bhavadharini RM. LuCoNet: A Convolutional Neural Network Model for Lung Cancer and Colon Cancer Prediction Using Histopathological Images. International Research Journal of Multidisciplinary Scope. 2024; 5(3):407-419. http://dx.doi.org/10.47857/irjms.2024.v05i03.0766
6. Yu Y, Feng T, Qiu H, Gu Y, Chen Q, Zuo C, Ma H. Simultaneous photoacoustic and ultrasound imaging: A review. Ultrasonics. 2024; 135:107277. https://doi.org/10.1016/j.ultras.2024.107277
7. Elyan E, Vuttipittayamongkol P, Johnston P, Martin K, McPherson K, Moreno-García CF, Jayne C, Sarker MM. Computer vision and machine learning for medical image analysis: recent advances, challenges, and way forward. Artificial Intelligence Surgery. 2022 Mar 22; 2(1):24-45.
8. Li X, Li M, Yan P, Li G, Jiang Y, Luo H, Yin S. Deep learning attention mechanism in medical image analysis: Basics and beyonds. International Journal of Network Dynamics and Intelligence. 2023; 2(1):93-116. https://doi.org/10.53941/ijndi0201006
9. Rebuffi SA, Gowal S, Calian DA, Stimberg F, Wiles O, Mann TA. Data augmentation can improve robustness. Advances in neural information processing systems. 2021 Dec 6;34:29935-48.
10. Goceri E. Medical image data augmentation: techniques, comparisons and interpretations. Artificial Intelligence Review. 2023 Nov; 56(11): 12561-605.
11. Elgendi M, Nasir MU, Tang Q, Smith D, Grenier JP, Batte C, Spieler B, Leslie WD, Menon C, Fletcher RR, Howard N. The effectiveness of image augmentation in deep learning networks for detecting COVID-19: A geometric transformation perspective. Frontiers in Medicine. 2021 Mar 1;8:629134.
12. Chlap P, Min H, Vandenberg N, Dowling J, Holloway L, Haworth A. A review of medical image data augmentation techniques for deep learning applications. Journal of medical imaging and radiation oncology. 2021 Aug;65(5):545-63.
13. Choudhari, A.: PCOS Detection Using Ultrasound Images. Kaggle. https://www.kaggle.com/datasets/anaghachoudhari/pcos-detection-using-ultrasound-images
14. Zhao J, Zhang Z, Chen B, Wang Z, Anandkumar A, Tian Y. Galore: Memory-efficient llm training by gradient low-rank projection. 2024 Mar 6. https://arxiv.org/pdf/2403.03507
15. Lafta NA. A Comprehensive Analysis of Keras: Enhancing Deep Learning Applications in Network Engineering. Babylonian Journal of Networking. 2023 Nov 26;2023:94-100.
16. Fan D, Yu H, Xu Z. PDSE: A Multiple Lesion Detector for CT Images Using PANet and Deformable Squeeze-and-Excitation Block. arXiv preprint arXiv:2506.03608. 2025. https://doi.org/10.48550/arXiv.2506.03608
17. Bhati D, Neha F, Amiruzzaman M, Guercio A, Shukla DK, Ward B. Neural network interpretability with layer-wise relevance propagation: novel techniques for neuron selection and visualization. In2025 IEEE 15th Annual Computing and Communication Workshop and Conference (CCWC). IEEE. 2025 Jan 6:00441-00447.
18. Venugopal V, Joseph J, Das MV, Nath MK. An EfficientNet-based modified sigmoid transform for enhancing dermatological macro-images of melanoma and nevi skin lesions. Computer Methods and Programs in Biomedicine. 2022 Jul 1;222: 106935.
19. Deepak G, Muralidharan C. Improved Tuberculosis Detection through Deep Learning. International Research Journal of Multidisciplinary Scope. 2024;5(2):540-548.
20. Lakhdari K, Saeed N. A new vision of a simple 1D Convolutional Neural Networks (1D-CNN) with Leaky-ReLU functions for ECG abnormalities classification. Intelligence-Based Medicine. 2022 Jan 1;6:100080.
21. Suha SA, Islam MN. An extended machine learning technique for polycystic ovary syndrome detection using ovary ultrasound image. Scientific Reports. 2022 Oct 12;12(1):17123.
22. Seo BK, Pisano ED, Kuzmiak CM, Koomen M, Pavic D, McLelland R, Lee Y, Cole EB, Mattingly D, Lee J. The positive predictive value for diagnosis of breast cancer: full-field digital mammography versus film-screen mammography in the diagnostic mammographic population. Academic radiology. 2006 Oct 1;13(10):1229-35.
23. Alhajim D, Ansari-Asl K, Akbarizadeh G, Soorki MN. Improved lung nodule segmentation with a squeeze excitation dilated attention based residual UNet.

Scientific Reports. 2025 Jan 30;15(1):3770.
24. Cheng H, Lian J, Jiao W. Enhanced MobileNet for skin cancer image classification with fused spatial channel attention mechanism. Scientific Reports. 2024 Nov 21;14(1):28850.
25. Gencer G, Gencer K. Advanced retinal disease detection from OCT images using a hybrid squeeze and excitation enhanced model. PloS one. 2025 Feb 7;20(2):e0318657.
26. Yuan J. Brain tumor image segmentation method using hybrid attention module and improved mask RCNN. Scientific Reports. 2024 Sep 4;14(1):20615.
27. Zhao Z, Chopra K, Zeng Z, Li X. Sea-net: Squeeze-and-excitation attention net for diabetic retinopathy grading. In2020 IEEE international conference on image processing (ICIP). IEEE. 2020 Oct 25:2496-2500.https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=9191345&casa_token=WWIPBJ6txZUAAAAA:zPYFoTZjwbZoQZRZ-DCfhjAsQWMB4KpILDilcEJR3WXSAqXI4i_cmodbZtZg2ds4JrXSw8h1qxAU&tag=1
28. Reyad M, Sarhan AM, Arafa M. A modified Adam algorithm for deep neural network optimization. Neural Computing and Applications. 2023 Aug; 35(23):17095-112.
29. Brauwers G, Frasincar F. A general survey on attention mechanisms in deep learning. IEEE Transactions on Knowledge and Data Engineering. 2021 Nov 9;35(4):3279-98.
30. Qin J, Wu J, Xiao X, Li L, Wang X. Activation modulation and recalibration scheme for weakly supervised semantic segmentation. In Proceedings of the AAAI conference on artificial intelligence 2022 Jun 28;36(2):2117-2125.
31. Ahad MT, Li Y, Song B, Bhuiyan T. Comparison of CNN-based deep learning architectures for rice diseases classification. Artificial Intelligence in Agriculture. 2023 Sep 1;9:22-35.
32. Ghosh A, Srinivasan K. EffiDenseGenOp: Ensemble Transfer Learning with Hyperparameter tuning using Genetic Algorithm Optimization for PCOS detection from Ultrasound Sonography Images. IEEE Access. 2025 Mar 24. https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=10937489
33. Kermanshahchi J, Reddy AJ, Xu J, Mehrok GK, Nausheen F. Development of a machine learning-based model for accurate detection and classification of polycystic ovary syndrome on pelvic ultrasound. Cureus. 2024 Jul 22;16(7):e65134.
34. Hosain AS, Mehedi MH, Kabir IE. 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); 27–28 Oct 2022; Kuala Lumpur, Malaysia. IEEE; 2022. p. 1–6. https://doi.org/10.1109/ICEET56468.2022.10007353
35. Poorani B, Khilar R. An innovative approach for PCO morphology segmentation using a novel MOT-SF technique. Discover Computing. 2024 Aug 19;27(1):27.