PolyCystic Ovary Syndrome Detection Using CapsuleNet and Synthetic Data

MSc Research Project
Data Analytics

Deepti Deepak Tiwari
Student ID: X21240302

School of Computing
National College of Ireland

Supervisor: Prashanth Nayak

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Deepti Deepak Tiwari
Student ID:	X21240302
Programme:	MSc in Data Analytics
Year:	2023
Module:	Data Analytics
Supervisor:	Prashanth Nayak
Submission Due Date:	14/08/2023
Project Title:	PolyCystic Ovary Syndrome Detection Using CapsuleNet and Synthetic Data
Word Count:	6094
Page Count:	20

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	Deepti Deepak Tiwari
Date:	18th September 2023

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

PolyCystic Ovary Syndrome Detection Using CapsuleNet and Synthetic Data

Deepti Deepak Tiwari
X21240302

Abstract

Today 6% to 12% of women who are of their reproductive age suffer from a common endocrine disorder known as PolyCystic Ovary Syndrome (PCOS) which is the most common cause of infertility in females. Out of these only a few get diagnosed and undergo proper treatment. The challenge is to detect PCOS on the base of infected and non-infected ultrasound of ovaries. This research proposed a framework which used deep learning methodologies and data augmentation for PCOS detection. The model implemented in this research is a capsule network to overcome the drawback of traditional CNN model with synthetic data later, the model was tuned to achieve better results. The dataset used here consists 3856 images of pelvic region ultrasound which can be categorized into two categories. Later the model is evaluated on accuracy, F1 score, recall, sensitivity, loss function training and test accuracy etc. This research demonstrates promising results that can be achieved to detect PCOS using Capsule network and synthetic data. This model can be used in medical field to detect PCOS at early stage which can mitigate future complications.

1 Introduction

National Institutes of Health (NIH) statistics show that women who are at the prevalence of PolyCystic ovary Syndrome are 65% of the total population of women in the world (Adiwijaya et al.; 2015). PCOS is mostly found in women who are their reproductive age. PCOS has gained its name because of the cyst formation that takes place on the ovaries the reason for the formation of cysts is due to ovulation failure which results in the formation of fluid sacs, these sacs consist of a hormone called androgen which cause the menstrual cycle disturbance (Nasim et al.; 2022). PCOS is a metabolic disorder that can be associated with multifactorial heterogeneous, complex genetic, endocrine and metabolic disorders which are caused due to chronic anovulation (Allahbadia and Merchant; 2011). Women who are diagnosed with PCOS suffer from other disorders like insulin resistance, hyperinsulinemia, abdominal obesity, diabetes 2 and also cardiovascular disease (Allahbadia and Merchant; 2011). In the previous day, PCOS was detected on the basis of hormonal imbalance (Hdaib et al.; 2022). The diagnosis of PCOS in women who belonged to the age group of 10-21 is difficult as the symptoms of PCOS such as gaining weight, acne, and hormonal imbalance overlap with the characteristic of puberty whereas the diagnosis is easy in women of age group above 25 years (Hdaib et al.; 2022).

Nowadays doctors advise getting a pelvic area ultrasound image to have a clear view of the follicle present in Ovaries. Apart from Rotterdam criteria are used to detect if women are suffering from PCOS or not which is based on the woman can be said that she is suffering from PCOS when a follicle of 12 or 2-9 mm is formed and the number of follicles is manually counted with the help of an ultrasound image, in order to have correct and thorough diagnosis knee and sharp medical skills are required but still, there is a scope of human error. Many studies were conducted to detect PCOS using different approaches used hormonal level but the hormonal level this hormonal disturbance can be caused due to external environmental factors as well later ultrasound images with deep learning were used to detect the infected ovaries, mostly CNN was used which were found to be the most fruitful and effective method (Hosain et al.; 2022). Saif et al. (2021) in his study stated that the spatial relationship of the input image is not taken into consideration at any instance by CNN. Hence all the images used in CNN should have the same orientation and focus area. In addition to that CNN cause a loss of image information in the max pooling layer which might be important. Moreover to make the model more accurate and flexible a large dataset is required. But collecting and creating a dataset related to one specific domain which has a similar image orientation is difficult.

In order to overcome the mentioned shortcoming of CNN mentioned above and its limitation this study proposes a capsule network model which considers the spatial relationship of images while working with them. The main aim of the research is to see how data augmentation improves capsule network in detecting PCOS.In addition to that the research also aims to classify the infected and non-infected ovaries and detect PCOS. The model proposed in this research would be used in the medical field to detect PCOS accurately and effectively. As a result, this model can be used for early detection of PCOS and start the treatment at an early stage which will avoid further complications and health issues that women commonly suffer while PCOS.

The research explains the study and methodology that were used so far in section 2 related work. The methodology used and design specification of the models and techniques used in this research is discussed in section 3 and section 4 respectively. The implementation of the research is explained in section 5 and evaluations conducted and their results obtained are discussed in section 6. Section 7 mentions the conclusion and future work.

1.1 Research Question

"How well a data augmentation improve CapsuleNet performance in PCOS detection? ".

2 Related Work

PolyCystic Ovary Syndrome (PCOS) detection has become an important research area in the medical field. Most women in this world suffer from PCOS knowingly and unknowingly as its symptoms are commonly ignored which makes the detection of PCOS very difficult. This makes it challenging to detect PCOS various study and research was conducted to detect PCOS using machine learning and deep learning methods.

2.1 PCOS detection using Machine learning and Deep-Learning Approaches

Bharati et al. (2020) conducted a study to detect PCOS using machine learning methodologies where they used a dataset which consists of 43 attributes which are related to hormone level fluctuations that are usually observed in patients suffering from PCOS. The study highlighted that the ratio of Follicle-stimulating hormone (FSH) and Luteinizing hormone (LH) plays an essential role in the detection of PCOS. The study used various classifiers such as s gradient boosting, random forest, logistic regression, and hybrid random forest and logistic regression (RFLR) from above mentioned classifier the study stated that random forest and logistic regression (RFLR) had better results as compared to others. The study implemented a univariate feature selection method to select the feature and avoid overfitting the model. However, the study did not consider that these hormone fluctuations can be caused due to many other factors which might not be related to PCOS. Hence, considering only hormone information to detect and classify PCOS is not sufficient.

Deshpande et al. (2014) study aims to detect PCOS by counting the number of follicles that were observed in ultrasound images of ovaries. This research pre-processed the pictures which include contrast enhancement and filtering to extract the necessary features and further used a support vector machine to classify the patients that are suffering from PCOs or not. The contrast enhancement technique used were histogram equalization, adaptive histogram equalization, and top-hat transformation used for feature extraction. The study showed promising results but its accuracy can be improved. In addition to that, the research does not mention the specificity and sensitivity of the model.

Dewi et al. (2018) performed the classification of ovaries with the help of ultrasound images. In this study, CNN was used for classification. The study explains about different pre-processing have been performed and extracted features from the image which can be used by CNN for classification. Hemming Net and Max Net were used to build the model as they can be easily combined with CNN. The WTA winner takes all approach was used where the neuron will compete among themselves and will produce only one winning neuron. The model achieved 80.84% of accuracy. However, the study did not specify how augmentation images helped the model improve its performance and accuracy. Moreover, the study does not mention any particular steps or procedure that was followed for image pre-processing.

Chitra et al. (2023) in their work demonstrated how deep-learning hybrid models can be used for the classification of PCOS using ultrasound images. This study used a univariate method for feature selection, scoring the features on pre-defined criteria and selecting the features with the highest score. This study pre-processed the image in which forward scaling was initially applied and resizing of the image was done in order to maintain the input image ratio. The study evaluated its model on precision, recall, sensitivity specificity and F1 score. The study used models like Alexnet with ideal and the study also highlighted that the accuracy of the model increases by changing the activation function to RELU moreover the study also highlighted how accuracy can be increased by just adding some dropout layers in VGG-16 in addition to that study highlighted how depth of the model plays a vital role in image classification by using Resnet 50 the study highlighted that efficiency was distorted by increasing the depth. The model proposed in this study used CNN which included VGG-16 and ResNet-50 to classify PCOS. The study proposed a state of an art model which was able to achieve an accuracy of 95%.

2.2 Capsule Network

Afshar, Plataniotis and Mohammadi (2020) conducted research in which they proposed a model to detect and classify the different types of tumors using Magnetic resonance Imaging. In their research, they proposed a model which used BootsCaps for tumor classification. This paper also mentioned the drawbacks and limitations of CNN with respect to poling layer when used for medical-related-image applications. The model proposed in this paper consists of a capsule network embedded in a boosting framework. They used segmented brain tumors instead of using the whole image with the help of coarse boundaries this area was concatenated with the output layer which formed a vector and utilized by the fully connected layer. They performed various experiments to evaluate and understand the working of BoostCap the 10 capsule network as the weak learner was used to compare with the proposed model the evaluation was done on the accuracy, sensitivity, specificity and area under the curve. The BoostCaps was able to achieve a decent accuracy of 92.45%, sensitivity of 96% specificity of 97% and AUC equal to 0.97. However, the accuracy can be improved by using a larger dataset and by image augmentation.

The work of Afshar, Heidarian, Naderkhani, Oikonomou, Plataniotis and Mohammadi (2020) explained how capsule networks can be used for the identification of COVID-19 cases from x-ray Images. They proposed a capsule model which consists of four convolution layers and three capsule layers. the loss function used in their work was modified in order to handle the imbalance in the dataset between negative and positive cases present. They performed various experiments to evaluate the model moreover they binarized the labels as positive or negative with the gaol to detect the positive case by examining the chest x-rays. In their work they used two different datasets one of the data sets was used to train the proposed model as they proceeded ahead they used this pre-trained model with new data set for the classification of positive and negative cases. The model was a state of the art which recorded an accuracy of 95% without pre-training whereas the pre-trained model accuracy was recorded at about 98%. The model has achieved a good result no doubt however the result might also be improved with the help of image augmentation and using the same dataset.

Afshar, Mohammadi and Plataniotis (2020) proposed a model called as BayesCap which can be explained as a combination of the Bayesian approach to classify the brain tumour classification using a capsule network. The architecture of the proposed model can break down into the first layer of convolution second layer of primary capsules and third layer of the main capsule and the last layer of the fully connected layer with softmax activation. The experiment performed in this research shows that when Monte-Carlo was implemented for 500 epochs the model achieved the accuracy of 68.3%. Whereas, the accuracy of the model was increased after using the Bayesian approach in order to improve the accuracy bootstrapping was done which increased the accuracy up to 73%. The Bayesian method was used to avoid the overfitting. However, the accuracy of the model can be increased by using different optimization techniques and a bigger datset.

2.3 Data Augmentation

Frid-Adar et al. (2018) conducted a study to explain how CNN performance can be improved using data augmentation. Where this research highlighted different ways of data augmentation and its benefits. The classic data augmentation is used to avoid shape deformation whereas the study also highlighted the number of operations like translation, rotation, scaling, and flipping that can be performed to achieve the classic data augmentation. They have also mentioned the second method for data augmentation which is Generative Adversarial Networks (GANs) to avoid over-fitting to the model. The model used in the research was Deep Convolutional GAN (DCGAN) to synthesize liver lesion images and tested whether the classification can be improved by testing various hypotheses. The result achieved by this experiment showed that the accuracy of the model was increased to 85.7%. The research has achieved a decent percentage of accuracy however the accuracy can be enhanced by using GAN with different models.

Mikołajczyk and Grochowski (2018) in their study demonstrate how data augmentation can be used for improving deep learning in image classification problems. The study mentions different types of data augmentation like traditional transformation, generative adversarial networks, texture transfer and other approaches. The study mentioned that traditional methods of data augmentation based on image transformation and colour modification are fast and easy methods that can be used to increase the training dataset. Moreover, the study highlights that the traditional method does not bring any new visuals to the image. Thereafter, the study also mentioned GANs and how they can be used in deep-learning methods in addition this study mentioned some limitations like Computational time, problems with counting, and trouble coordinating the global structure of the GAN method. However, the study did not mention how data augmentation can be helpful to improve the efficiency of the deep learning method.

Khosla and Saini (2020) compared different data augmentation techniques in order to enhance the performance of deep learning models. The different data augmentation techniques discussed were geometric transformation Flipping, Colour Space, Cropping, Rotation, Translation, and Noise injection. Moreover, this paper also discusses Data Augmentation based on Oversampling and different methods used like mixing images, feature-space augmentation and Generative Adversarial Networks and how they can be implemented to increase efficiency. Moreover, the paper also discusses different techniques like transfer learning, drop-out and batch normalization and pre-training to be useful to overcome overfitting. This study provided an overview of different techniques that can be used to increase the training dataset and how they can be helpful in increasing the efficiency of the deep learning model. However, the study did not specify which data augmentation method can be more effective with respect to different domains.

Salehinejad et al. (2018) in their work discussed how the radial transformation method can be used for image data augmentation. The radial transformation uses spatial and polar coordination, generating the new image representation. The study highlighted the advantage that radial change brings is the ability to up-sample the pixels. The study has evaluated how well radial images can effectively use to increase the performance of the deep learning model. The study concludes that the image generated by radial transform performed better and achieved good results.

3 Methodology

Two well-known and widely used methodologies like e Cross-Industry Standard Process of Data Mining (CRISP-DM) and Knowledge Discovery in Databases (KDD) are used for data understanding and extracting useful information. This research have used the KDD methodology to understand and extract valuable information from the dataset and used it to complete this research and achieve the desired results. There are six steps that are followed as mentioned below:

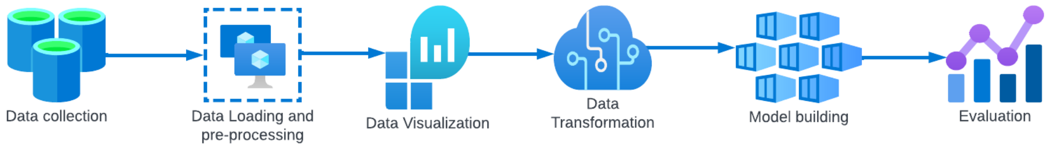


Figure 1: Knowledge Discovery in Databases (KDD)

1. Data Selection.
   This is the basic and first step in KDD. where purpose of the research was understood clearly and an appropriate dataset was selected.

2. Data loading
   The data need to be stored properly in order to retrieve the data while implementation.

3. Data Visualization
   In this step the data is visualized to understand the data better. These steps also help to understand the distribution of the data and flag if there are any class imbalances.

4. Data Transformation
   This step involves transforming the data into the right format which is more suitable and compatible with the model.

5. Data Mining
   The data mining is the core stage as in this phase we need to choose the right and correct model that can be implemented to meet the objective of the research. This phase helps us to understand the trend, and pattern of the transformed data and which techniques from classification, clustering, regression and prediction are better fit for the data.

6. Evaluation
   The result obtained from the model is evaluated to extract valuable insight. On the basis of the evaluation and conclusion drawn and model can be continuously improved on feedback.

The proposed system uses "PCOS detection using ultrasound images dataset". The dataset obtained from Kaggle consists of train and test folders these folders consist of labelled as infected and non-infected images. The data is saved in google drive which is further used along with the model proposed. Then these images are loaded, pre-processed and split into train and test data. Then one-hot encoding is performed for the target labels. Then, the Capsule Network model is defined layer by layer and later model was compiled. The model is then evaluated on the basis of sensitivity, specificity, F1 score and accuracy score. Then data augmentation is performed in which the images are flipped by 30 degrees, 60 degrees and 90 degrees and then these images are used with the pre-trained model of base Capsule Network. The data formed after augmentation is known as synthetic data. The accuracy of the model is checked by using the synthetic data and original data ratio in different ratios 1:1, 1:2 and 1:3. Proceeding ahead the model is multi-staged fined tuned and evaluated on original and synthetic data. Moreover, this study consists of several experiments to make the designed model more compatible with real-world environments and scenarios. The methodology of this research can be summarised in the below flow chart 2.

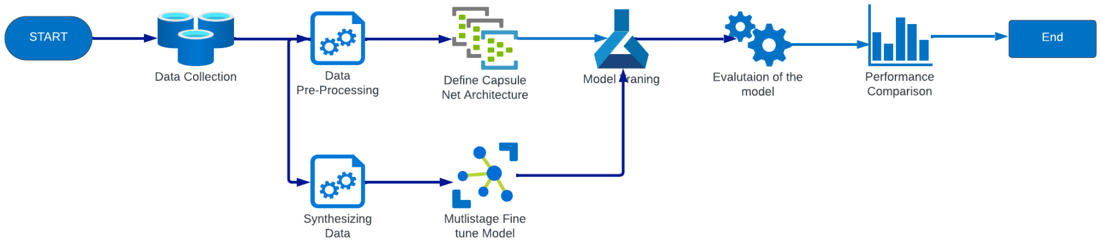


Figure 2: Research workflow

3.1 Data Collection and Understanding

This research used the dataset "PCOS detection using ultrasound images" which can be easily retrieved from Kaggle [1]. The dataset consists of 3856 files which are around 132 MB. The data is then divided into testing, training and validation. Where both the testing and training folders comprised of 2000 images approximately. There are two classes in this folder infected and not infected.

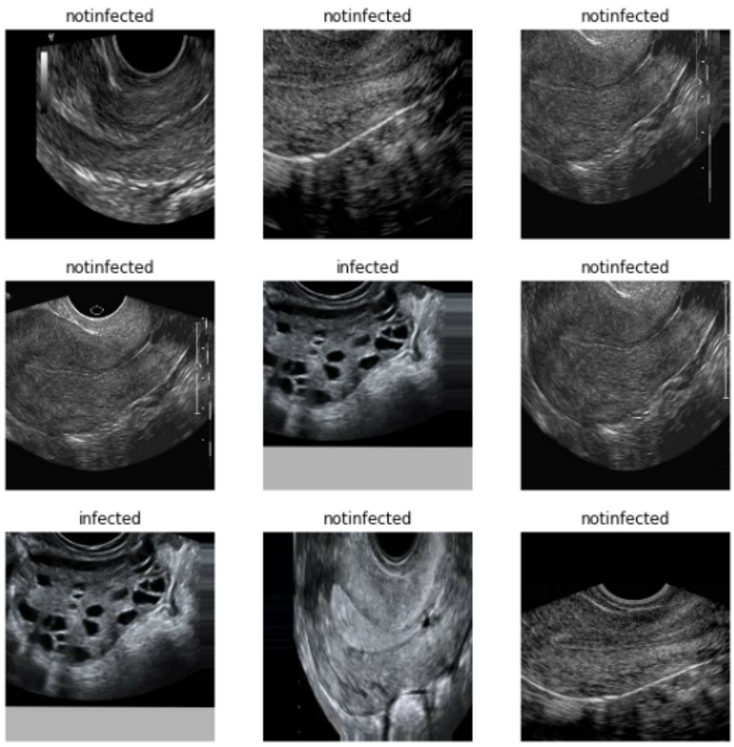
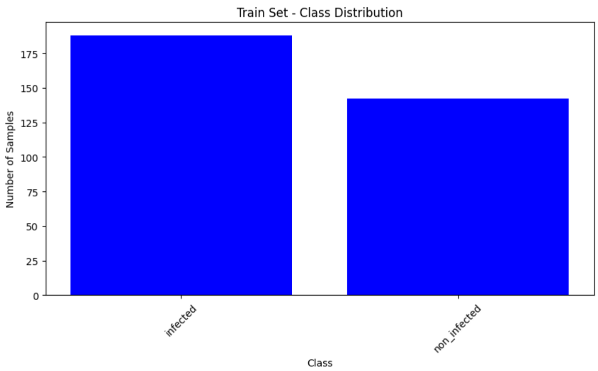


Figure 3: Sample Data

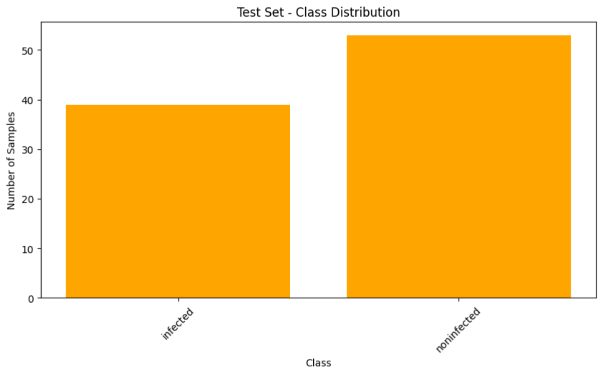
[1]Dataset link: https://www.kaggle.com/datasets/anaghachoudhari/pcos-detection-using-ultrasound-images.

3.2 Data Pre-Processing

In this research, the dataset used can be easily retrieved from Kaggle [2]. This dataset was divided into training and testing. Then moving ahead data was split into 8:2 into train and validation. This dataset was fairly very large but after going through the dataset it was found that there were large amounts of similar images present in both test and train data which resulted in overfitting of the model. In order to avoid overfitting, similar images were deleted from both folders and only the unique ones were Kept. This resulted in the reduction of dataset size. In order to overcome this data augmentation was performed which helped to improve the accuracy of the model.



(a) Class Distribution for Train Data



(b) Class Distribution for Test Data

Figure 4: Data Distribution

In order to avoid the biasness of the model a class check was performed to verify if all the classes are balanced. From the figure 4 it can clearly observe that there is class imbalance in the data. The infected images were more in train data as compared to test data on the other hand the non-infected images were less in train data in comparison to test data. Hence, classes should be balanced before it is used along with the proposed model.

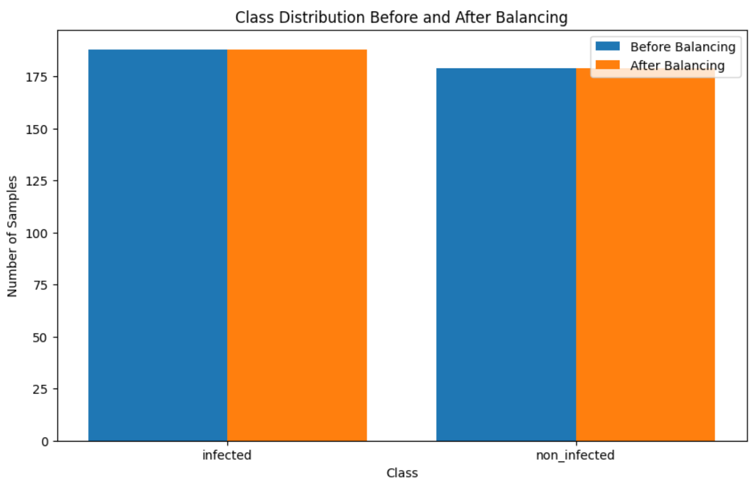


Figure 5: Balanced Data

[2]https://www.kaggle.com/datasets/anaghachoudhari/pcos-detection-using-ultrasound-images

In order to balance the class random sampling was performed. Random sampling is a common method used to balance image data. This methods add random images by creating a subset of images to each category in order to represent them more equally. Hence, data was balanced after performing Random Sampling as shown in figure 5.

3.2.1 Data Augmentation

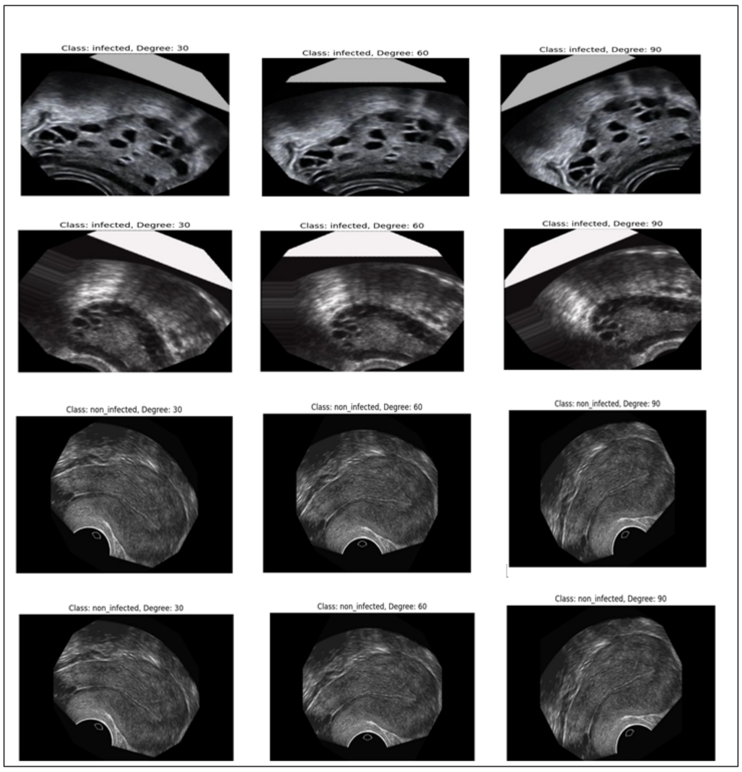


Figure 6: Data Augmentation

Data augmentation can be defined as the artificially generating the data by modifying and copy the existing data present [3]. Data augmentation is used to overcome the problem related to the limited dataset (Alomar et al.; 2023). Data augmentation helps in the expansion of a training dataset. There are different techniques like flipping, rotation, translation, Scaling, brightness and contrast adjustment, noise injection, crop and resizing of images. used to generate the synthetic image. Alomar et al. (2023) in their discussion concluded that image generated using the random local rotation are more effective as it does not manipulates add or add pixel value. Proceeding ahead image was rotated at different angles to create synthetic images and observe how the model performed on it. The existing data was augmented using ImageDataGenerator function which defined different parameters that have been used on existing training data. The parameters that were considered are rotation_range, width_shift_range, height_shift_range and horizontal_flip. Where rotation range specifies the random range of rotation in degree, width_shift_range, height_shift_range defined random shifts respectively and in order to enable the random horizontal flipping the parameter horizontal_flip was defined. 'datagen.flow' method was used which uses the original data and generates the augmented images as per the defined

[3]https://www.datacamp.com/tutorial/complete-guide-data-augmentation

parameter. During each epoch defined ImageDataGenerator performs the augmentation of the existing train dataset which generates new synthetic images with variation. Further to justify the research question the traditional data rotation technique was used to generate the synthetic data at specific angles and this augmentation was done on existing train data and later new synthetic datasets were used in research. To summarize in short the existing data was augmented by defining a specific range for rotation, width, height and horizontal flip.

4 Design Specification

4.1 Model Building

As discovered in the literature review, deep-learning models are usually used for classification tasks. Most of the literature review showed that CNN performs well on image datasets when used for classification purposes. But in work of Saif et al. (2021) showed that whenever there is a change in the image the CNN needs to be trained again on the new data acquired which might be time-consuming. So in this research, Capsule-Network is implemented as it considers the spatial relations between images at any instant.

4.1.1 Capsule Network

Capsule Networks are a type of neural network that was used to overcome the limitations of traditional Convolution Neural Networks (CNNs) in handling variations and hierarchical relationships in image data. They have recently gained attention due to their property which encrypts the relationship between entities like scales, location, pose and orientation (Haq et al.; 2023a). Moreover, Haq et al. (2023b) through his study stated that capsule networks store high vector space information which includes image colour, texture, position, orientation etc. Capsule Networks use Dynamic routing to communicate between the neurons. The basic principle that the routing algorithm work on is it refines the output gained from the upper level of the capsule network based on the input obtained from the lower capsule layers.

The Capsule network Architecture consists of Primary capsules which are responsible for identifying the local features in input data. The second layer of the model consists of digit capsules which are mainly responsible for extracting more high-level features from the input data. Each digit capsule represents a specific class or object. In this research, the higher capsule represents the classes of infected and non-infected ovaries.

Capsule Network Architecture used in this study consists of 256 convolution layers with a 9x9 filter and stride of 1 which is the first layer. The second layer i.e. primary capsule network is formed by a 9x9 filter and 256 convolution layer and stride of 2 whereas, the third layer of this model performs routing by agreement process. The last and final layer of the model is a fully connected layer with softmax activation as shown in figure 7. Moreover, Adam optimizer and categorical cross-entropy along with metrics accuracy are used while compiling the model.

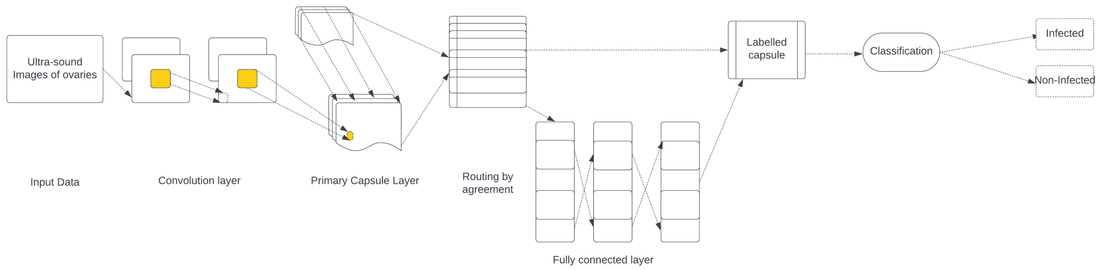


Figure 7: Capsule Network Architecture

4.1.2 Multistage fine-tuning of Capsule Network

This research has performed multistage fine-tuning on the above model 4.1.1 to enhance the performance and accuracy of the model proposed. First, the base capsule network was trained on the original dataset which helped the model to learn basic features that were relevant across different tasks. Fine-tuning can be defined as where weights and parameters are adjusted at different values and observation model performance is made. Whereas, in this research work lower layer of the capsule network was frozen which preserved the previous knowledge gained in the training stage and allows higher-level capsules to adapt in order to perform a target-specific task.

Proceeding ahead, the entire model was trained with lower learning. Here, a lower learning rate is applied to retain previously learned features. The learning rate defined in this study is 0.0001.

4.1.3 Creating Synthetic dataset and Monitoring

This study consists of the synthetic data which was generated to test the performance of the model. This synthetic data was generated by using traditional data augmentation which implemented the rotation method to rotate the original images at different angles, from the literature review it was concluded that this was one of the effective methods. This research has used only the first quarter which is angles between 0 to 90 degrees to rotate the image. The angles used to generate the synthetic data was 30, 60 and 90 degrees which can be justified as these angles showed major difference and improvement in model performance.

5 Implementation

This research used three layered Capsule Network Architecture, as described in the previous section 4.1.1 along with data augmentation. The subsection gives a brief idea about the model, dataset, and experiments performed in order to evaluate the models.

5.1 Environmental Setup

In order to implement the proposed model google colab was used. The hardware specifications used were as followed 16GB RAM, a 458 GB hard drive, and a GPU NVIDIA GEFORCE. The dataset was loaded to Google Drive. The Google Colab is a powerful

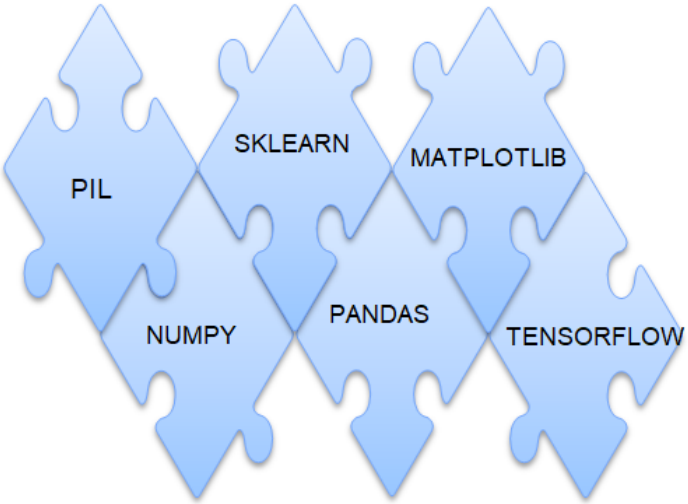


Figure 8: Libraries

tool as it takes less time to process any size of the dataset. The Google Colab used Python 3 Runtime with T4 GPU hardware accelerator. Coding for this research was carried out with the help of Python language. Python language has recently got its increasing popularity in data analytics and data science fields because of its flexibility and the wide range of libraries that it brings along with it, which makes the building and execution of any machine, deep-learning model very easy. In this research, various common libraries like numpy, PIL, matplotlib, Sklearn, ImageDataGenerator, pandas, sklearn and TensorFlow were used as shown in figure 8.

5.2 Data Processing

To make data more readable and useable with the model data was processed as described in this section. The dataset was divided into test and train by default. In order to use these images along with the model the pre-processing of the image was done. The images were cleaned by removing the noise present in them, in addition to that images were transferred to grayscale images, to make the processing of data easier the images were reshaped and the same size of images for train and test was used later, the images in both train and test were randomly flipped and necessary shifts were performed. At last the normalization of data was done. Later, label encoding is carried out to classify the correct number of target classes this is achieved by using the function LabelEncoder.

5.3 Data Augmentation and Model Building

The dataset used consists of ultrasound images of infected ovaries and healthy ovaries. This dataset was read from google drive and then used with the model. The model was built with the help of Keras where the model was defined as per the design specification described in 4.1.1 Conv2D function was used to define the convolutional layer with activation function as relu and input shape whereas Flatten and Dense function where used to define third and last layer of the Capsule network respectively. Later the model was called by passing input shape and number of classes and compiled using Adam optimizer and losses were calculated using categorical cross-entropy. The model was tested for 10 epochs and 32 batch size. Then the model was evaluated against the validation data and performance was recorded.

As we further proceed in this research Data Augmentation was performed on the original image. Where both train and test data were rotated at a specific angle as described in section 4.1.3. A function rotate_image_def was defined to rotate the data images at the specific angles mentioned. The base Capsule Network model which was trained on original data was saved and the same model was loaded and then trained on augmented generated data. Later different ratios of the original image to synthetic images were used to evaluate the model performance. The different ratio of original to synthetic data was used to see how the performance of the model after it was trained with augmented data.

This research also highlights how the model was multistage fine-tuned further. Further, the fined tuned model was used on two different datasets one was the original dataset and the synthetic data that was created in the previous stage of this research. The model was fine-tuned as described in section 4.1.2. the layers were frozen by initiating them to false at the time of training the data later whole model was trained with a small learning rate that is 0.0001.

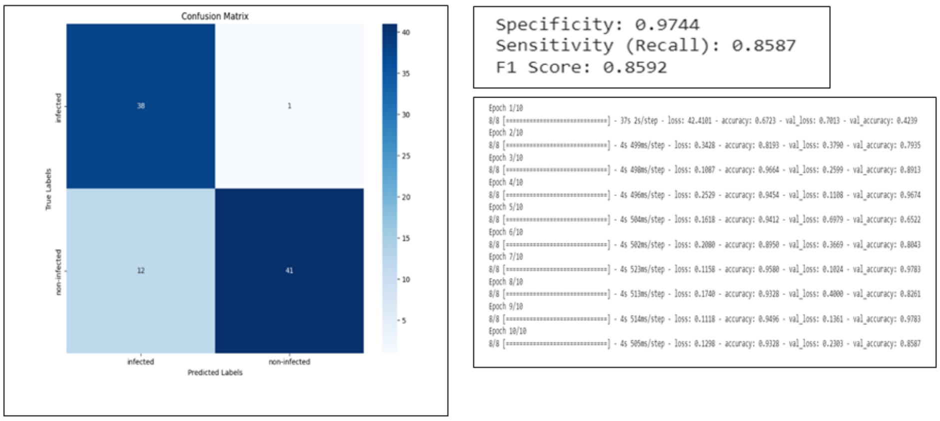
In order to carry out the experiment smoothly and effectively the path for the test and train data directory were set. Input image size was defined which can be later used by the model while execution. Then research evaluated the model on the basis of accuracy, AUC, F1 score Confusion Matrix, sensitivity and specificity. The above parameters are used for evaluation as this research is related to the medical field and in the medical field true-negatives and false-positive play a more important and significant role in comparison to the accuracy of the model.

6 Evaluation

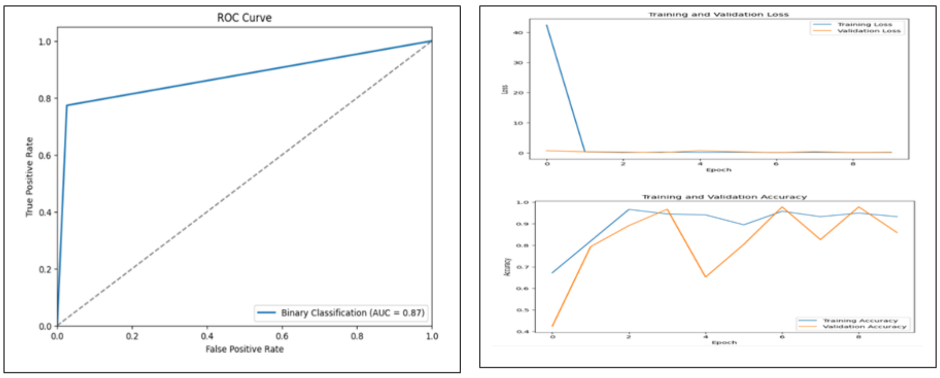
The main objective of this research is to how data augmentation can help Capsule network to classify the PCOS. In order to study that and for firm understanding a set of different experiments were performed and the results were observed. The evaluation of this experiment was done with the help of accuracy, AUC, F1 score Confusion Matrix, sensitivity and specificity. The results achieved from experiments will help us to understand how effectively data augmentation helps the Capsule network to perform the binary classification. Moreover, the research has also evaluated the multistage fine-tune model and its performance with synthetic data generated. In addition to that models are trained and evaluated at different epochs, dropout rates and tunning different layers of the model. Test data is used for evaluating the model,

6.1 Experiment 1: Classification with Capsule Network

The goal of the first experiment was to study how effectively binary classification is performed by capsule networks on image data. In this experiment, the base capsule network with three layers is used. These three layers can be breakdown as a convolution layer, a primary capsule layer Routing layer and a fully connected layer of the model. The base model used 10 epochs and 32 batch size and input image as 128x128. The accuracy recorded for the test dataset was 85.87% and the loss was about 23.03% as losses were decreasing with increasing epochs. Where sensitivity recorded for this experiment was 85.87%, specificity was 97.44% and F1 Score was 85.92%. From the confusion matrix,



(a) Confusion Matrix and execution details



(b) Plots for experiment 1

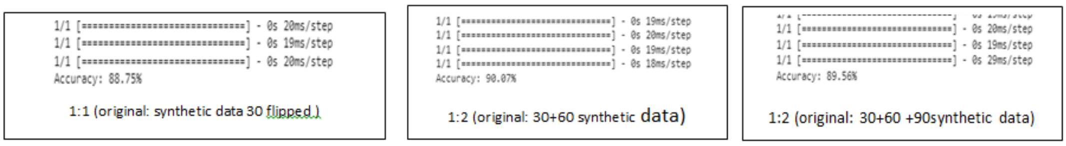
Figure 9: Evaluation for experiment 1

one can observe that there was 1 True-Negative and there were 12 False-positive on the other hand True-Positive and False-Negative were 38 and 41 respectively.

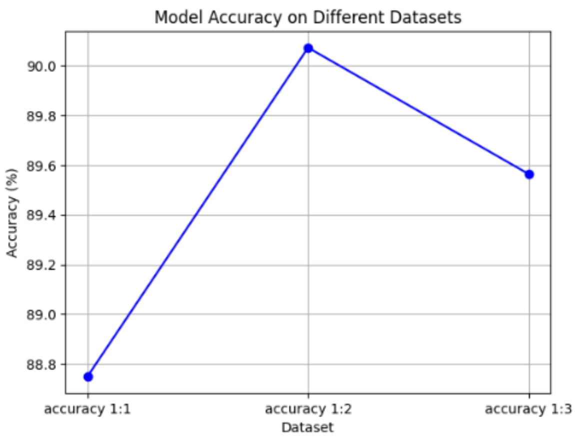
6.2 Experiment 2: Capsule Network on Augmented images

The second experiment's aim was to study the performance of the capsule network on synthetic data. In order to create the synthetic data the original image was rotated at different angles like 30, 60 and 90 degrees. Then a capsule network was used to perform the classification of this synthetic data. The accuracy recorded for synthetic data produced by flipping the data by 30 degrees was recorded as 42.39%, The sensitivity and specificity recorded were about 47% and 53% respectively throughout the three scenarios. Whereas, the accuracy recorded for 60 degrees flip image and 90 degrees flip synthetic data was the same as 30-degree flip synthetic data. Here a clear observation can be made that the Capsule network did not perform well when only synthetic data was used.

6.3 Experiment 3: Combining original and synthetic data into different proportions to observe the model performance



(a) Confusion Matrix and execution details



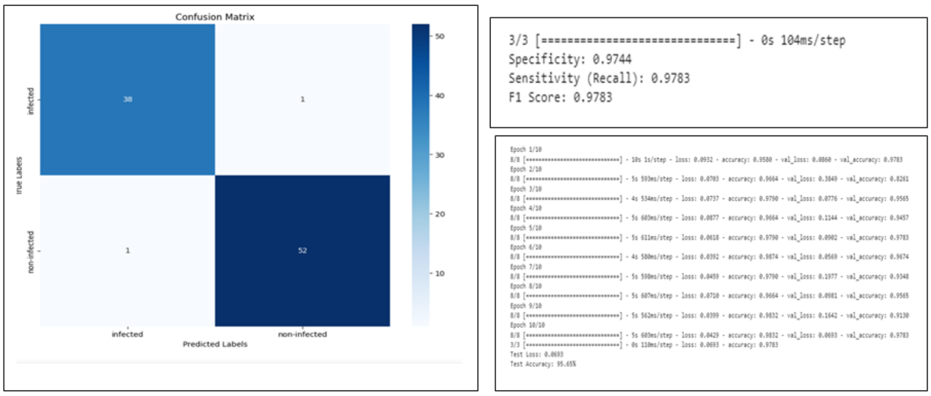
(b) Plots for experiment 1

Figure 10: Evaluation for experiment 3

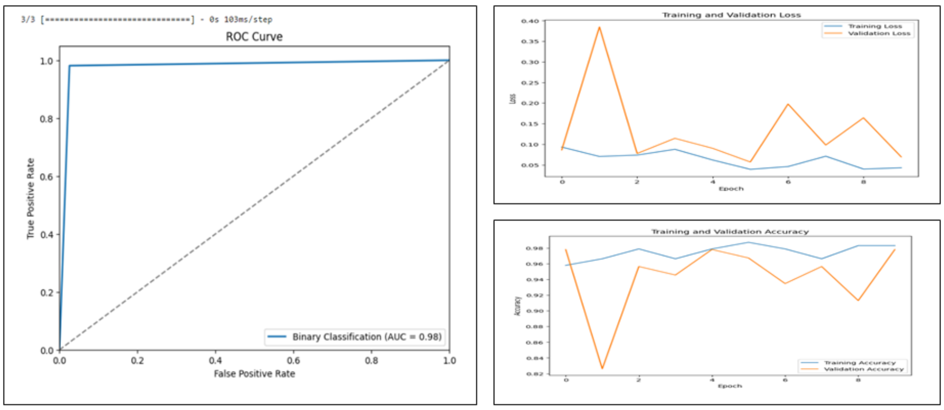
The third experiment of this research was performed with the objective was conducted to see the performance of the pre-trained capsule network. This experiment was divided into three parts where a data set was created by combining original images and synthetic data. The model recorded an accuracy of 88.75% for the dataset created in the ratio of 1:1 which consist of the original image and 30-degree flipped synthetic images. Whereas the accuracy of the model increased with increasing ratio of synthetic data the accuracy recorded for 1:2 which can be explained as the ratio of an original image to 30 and 60 degrees of synthetic data was about 90% and the accuracy recorded for 1:3 can be explained as ratio of the original image to 30, 60 and 90 degrees flipped synthetic data was recorded about 89.5%. From this experiment, the study tried to explain that the model performed well with the mixture of original and synthetic data when used in different portions. Experiment three can be concluded as the best results are achieved when the dataset is combined in the ratio of 1:2 as it has less noise in its image data and is more suitable for the image to process.

6.4 Experiment 4: Multistage fine-tuning the capsule network

In the fourth experiment, the aim was to study the behaviour of multistage fine-tune model of the capsule network. The tuned model used Adam optimizer and categorical



(a) Confusion Matrix and execution details

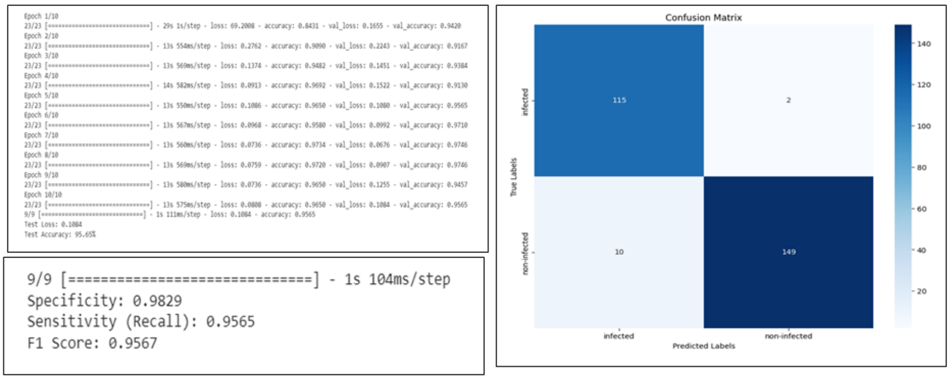


(b) Plots for experiment 4

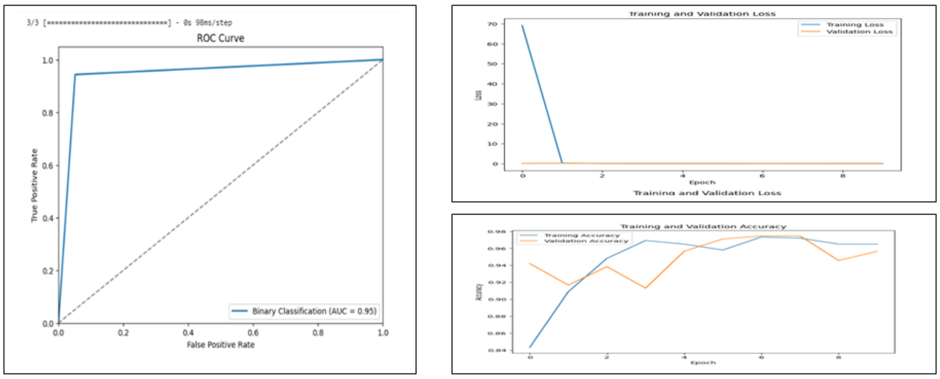
Figure 11: Evaluation for experiment 4 with original data

cross-entropy to calculate the loss, along with 32 batch sizes and up to 5 epochs. The accuracy recorded was about 42.39 % throughout the five epochs. After performing image augmentation on the rotated images and training the model again the accuracy of the model was improved to 95.65% and test loss recorded was 0.1084 where as specificity and sensitivity were recorded at about 98.29% and 95.65% respectively and f1 score was equal to 95.67%. When the confusion matrix was plotted the total True-Negative was 2 and there were 10 False-positive on the other hand True-Positive and False-Negative were 115 and 149 respectively. The ROC curve plotted is as shown in the below figure with AUC= 0.96.

When the same multistaged fine-tunned model was used with the original dataset the results recorded were as follows accuracy of 95.65% loss equal to 0.069 Specificity 97.44% Sensitivity (Recall) equal to 97.83% F1 Score equal to 97.83%, confusion matrix can be described as True-Negative was 1 and there was 1 False-positive on the other hand True-Positive and False-Negative were 38 and 52 respectively. The ROC curve plotted with AUC= 0.98.



(a) Confusion Matrix and execution details



(b) Plots for experiment 4

Figure 12: Evaluation for experiment 4 with original data

6.5 Discussion

Table 1: Accuracy.

Model	Dataset	Acurracy
Capsule Network	Original dataset	85.57%
Capsule Network	Synthetic data (1:1)	88.75
Capsule Network	Synthetic data (1:2)	90%
Capsule Network	Synthetic data (1:3)	89.5%
Fine Tune Model	Original dataset	94%
Fine Tune Model	Synthetic dataset	95.65%

This research was performed to see how data augmentation helps the capsule network to enhance its performance. The result achieved from this research and experiments is summarised below.

From the table 1 above we can see that the model performance has been improved when synthetic data started to add. In order to enhance the model performance multistage model tunning was performed which shows improvement in the model performance. The accuracy of the model was improved from 85.57% to 94% when the model was fine-tuned and used along with the original dataset. Whereas, the accuracy of the model was improved from 89.57% to 95.65% when the model was fine-tuned and used along with the synthetic dataset.

Table 2: Sensitivity and Specificity.

Model	**Dataset**	**Sensitivity**	**Specificity**
Capsule Network	Original dataset	85.87%	97.44%
Capsule Network	Synthetic data (1:1)	87.89%	97.44%
Fine Tune Model	Original dataset	97.83%	97.44%
Fine Tune Model	Synthetic dataset	95.65%	98.29%

Table 2 shows how data argumentation increases the sensitivity of the model whereas specificity remained the same. On the other hand, it can be observed that when the model is fined tune and used with original and synthetic data the the sensitivity decreased whereas the specificity of the model was increased.

Moreover, from experiment three 6.3 a conclusion can be drawn that good accuracy results were achieved when data was in 1:2 portion used with the base model but increase the ratio of the original image to the synthetic image resulted in deterioration of the accuracy when used along with based Capsule network model. Whereas, from experiment four 6.4 conclusion drawn is that when original data with a fined-tuned capsule network is used in comparison to synthetic data the synthetic data with fine tune model gives better result. Hence a it can be stated that data augmentation does help in the performance enhancement of the capsule network.

7 Conclusion and Future Work

The primary objective of the research was to verify if data augmentation helps enhance the performance of capsule networks in the detection of PCOS. From the experiments conducted so far in this research, it can be concluded that sure data augmentation enhances the model performance but, in some scenarios like experiment three the performance starts decreasing after a certain ratio. Whereas a boost in performance can be seen after fine-tuning the model. Hence, this model can be implemented in the real world to detect PCOS at an early stage with the help of ultrasound images which can avoid the human error of counting the follicle and misdiagnosis.

The research required a large computational time future work can include how this computational time can be reduced using Bayesian optimization or any other optimization techniques. DALL-E technique can be used to generate the synthetic data can help to overcome the challenges of data limitations due to privacy concerns. It can also be used to generate images of the ovary by the anatomical and biological descriptions which can describe different textual and anatomical structures of the Follicle. Even rare and abnormal cases of ultrasound images of PCOS can also be generated with the help of

DALL-E which are often neglected while creating a real-world dataset this will introduce diversity in the training dataset which can help to enhance the model performance.

Acknowledgement

Firstly I would like to thank my supervisor Mr Prashanth Nayak for his constant support and valuable guidance during this research. I would also like to thank my parents, friends, classmates and NCI College for their support and encouragement.

References

Adiwijaya, K., Purnama, B., Hasyim, A., Septiani, D., Wisesty, U. N. and Astuti, W. (2015). Follicle detection on the usg images to support determination of polycystic ovary syndrome, *Journal of Physics: Conference Series* **622**: 012027.

Afshar, P., Heidarian, S., Naderkhani, F., Oikonomou, A., Plataniotis, K. and Mohammadi, A. (2020). Covid-caps: A capsule network-based framework for identification of covid-19 cases from x-ray images, *Pattern recognition letters* **138**.

Afshar, P., Mohammadi, A. and Plataniotis, K. (2020). Bayescap: A bayesian approach to brain tumor classification using capsule networks, *IEEE Signal Processing Letters* **27**: 2024–2028.

Afshar, P., Plataniotis, K. and Mohammadi, A. (2020). Boostcaps: A boosted capsule network for brain tumor classification, Vol. 2020, pp. 1075–1079.

Allahbadia, G. N. and Merchant, R. (2011). Polycystic ovary syndrome and impact on health, *Middle East Fertility Society Journal* **16**(1): 19–37.
URL: *https://www.sciencedirect.com/science/article/pii/S111056901000141X*

Alomar, K., Aysel, H. I. and Cai, X. (2023). Data augmentation in classification and segmentation: A survey and new strategies, *Journal of Imaging* **9**(2).
URL: *https://www.mdpi.com/2313-433X/9/2/46*

Bharati, Subrato, Podder, Prajoy, Mondal, H. and Rubaiyat, M. (2020). Diagnosis of polycystic ovary syndrome using machine learning algorithms, *2020 IEEE Region 10 Symposium (TENSYMP)*, pp. 1486–1489.

Chitra, P., Srilatha, K., Sumathi, M., Jayasudha, V., F., Bernatin, T. and Jagadeesh, M. (2023). Classification of ultrasound pcos image using deep learning based hybrid models, *2023 Second International Conference on Electronics and Renewable Systems (ICEARS)*, pp. 1389–1394.

Deshpande, S., S., Wakankar and Asmita (2014). Automated detection of polycystic ovarian syndrome using follicle recognition, *2014 IEEE International Conference on Advanced Communications, Control and Computing Technologies*, pp. 1341–1346.

Dewi, M, R., Adiwijaya, Wisesty, N, U. and Jondri (2018). Classification of polycystic ovary based on ultrasound images using competitive neural network, *Journal of Physics Conference Series* **971**(1): 012005. https://doi.org/10.1088/1742-6596/971/1/012005.
URL: *https://app.dimensions.ai/details/publication/pub.1103305870*

Frid-Adar, M., Diamant, I., Klang, E., Amitai, M., Goldberger, J. and Greenspan, H. (2018). Gan-based synthetic medical image augmentation for increased cnn performance in liver lesion classification, *Neurocomputing* **321**: 321–331.
URL: *https://www.sciencedirect.com/science/article/pii/S0925231218310749*

Haq, M. U., Sethi, M. and Rehman, A. (2023a). Capsule network with its limitation, modification, and applications—a survey, *Machine Learning and Knowledge Extraction* **5**: 891–921.

Haq, M. U., Sethi, M. and Rehman, A. (2023b). Capsule network with its limitation, modification, and applications—a survey, *Machine Learning and Knowledge Extraction* **5**: 891–921.

Hdaib, D., Almajali, N., Alquran, H., Mustafa, W. A., Al-Azzawi, W. and Alkhayyat, A. (2022). Detection of polycystic ovary syndrome (pcos) using machine learning algorithms, *2022 5th International Conference on Engineering Technology and its Applications (IICETA)*, IEEE, pp. 532–536.

Hosain, A. S., Mehedi, M. H. K. and Kabir, I. E. (2022). Pconet: A convolutional neural network architecture to detect polycystic ovary syndrome (pcos) from ovarian ultrasound images, pp. 1–6.

Khosla, C. and Saini, B. S. (2020). Enhancing performance of deep learning models with different data augmentation techniques: A survey, *2020 International Conference on Intelligent Engineering and Management (ICIEM)*, pp. 79–85.

Mikołajczyk, A. and Grochowski, M. (2018). Data augmentation for improving deep learning in image classification problem, *2018 International Interdisciplinary PhD Workshop (IIPhDW)*, pp. 117–122.

Nasim, S., Almutairi, M., Munir, K., Raza, A. and Younas, F. (2022). A novel approach for polycystic ovary syndrome prediction using machine learning in bioinformatics, *IEEE Access* pp. 1–1.

Saif, A., Imtiaz, T., Rifat, S., Shahnaz, C., Ahmad, M. O. and Zhu, W.-P. (2021). Capscovnet: A modified capsule network to diagnose covid-19 from multimodal medical imaging, *IEEE Transactions on Artificial Intelligence* **PP**: 1–1.

Salehinejad, H., Valaee, S., Dowdell, T. and Barfett, J. (2018). Image augmentation using radial transform for training deep neural networks, *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 3016–3020.