Detection of Polycystic Ovarian Syndrome using Convolutional Neural Network in conjunction with Transfer Learning Models

MSc Research Project
Data Analytics

Nithya Sathiadhas Puvaneswari
Student ID: x21116270

School of Computing
National College of Ireland

Supervisor: Bharat Agarwal

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Nithya Sathiadhas Puvaneswari
Student ID:	x21116270
Programme:	Data Analytics
Year:	2022
Module:	MSc Research Project
Supervisor:	Bharat Agarwal
Submission Due Date:	15/12/2022
Project Title:	Detection of Polycystic Ovarian Syndrome using Convolutional Neural Network in conjunction with Transfer Learning Models
Word Count:	7842
Page Count:	23

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:	
Date:	14th December 2022

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):	

Detection of Polycystic Ovarian Syndrome using Convolutional Neural Network in conjunction with Transfer Learning Models

Nithya Sathiadhas Puvaneswari
x21116270

Abstract

Polycystic Ovarian Syndrome, commonly abbreviated as PCOS, is a medical ailment found in women aged between 17 to 40 years. It is a reproductive disorder that causes infertility in women during their childbearing age, may cause an irregular menstrual cycle, and may lead to multiple ovarian diseases. This medical disorder is one of the most concerning diseases in their reproductive age and may lead to long-term complications. Considering the disorder's uncertainty, the treatment remains undiagnosed for extended periods. Hence, a method is needed to diagnose and detect the presence of a growing number of follicles in the ovary so that the disease can be detected at an early stage. With the evolution of technology, deep learning methods have been immensely accepted for medical diagnosis, and varying networks with multiple layers have been adopted to generate desired results using image classification. Despite the fact that the implementation of deep learning algorithms leads to the generation of image classification with high levels of accuracy, it bears the limitation of consuming large network bandwidths, including computational complexity and execution time. Hence, for the implementation of the same, the thesis puts forward a model that could automate diagnosing the disease using ultrasound images of the ovaries. For this purpose, the conceptual theory of Convolutional Neural Network (CNN) is used in conjunction with transfer learning models: EfficientNet-B2 and ResNet-50. In addition to the usage of the dataset, the data augmentation process has also been implemented due to the limited number of available data. On conduction of the experiment, it was observed that the CNN model generated the highest accuracy of 91% compared to EfficientNet-B2 and ResNet-50.

KeyWords: CNN, Data Augmentation, Deep learning, EfficientNet-B2, PCOS, ResNet-50

1 Introduction

One of the most commonly occurring hormonal disorders observed in women is witnessed to be Polycystic Ovary Syndrome (PCOS). The hormonal imbalance caused due to PCOS tends to affect women of the age group 18-44 years (Morang et al.; 2019). The disease was initially identified in 1935 by Leventhal and Stein, wherein they defined the occurrence of PCOS due to the presence of enlarged cysts in the ovaries that might further lead to infertility in women. One of this disorder's significant symptoms is chronic anovulation that creates multiple cysts on the outer edges of the ovary. Furthermore, it causes a hormonal imbalance within the body, resulting in varying Luteinizing Hormone (LH) and Follicle Stimulating Hormone (FSH), which are essential for the growth of the egg inside the female body. In addition to the enlargement of cysts and egg maturation, the human body's entire reproductive system is affected, which might impact the development of follicles in the ovary. This prolonged development of follicles leads to an irregular menstrual cycle followed by unusual weight gain and unwanted hair growth on the face. Early detection of PCOS helps detect other reproductive diseases, such as infertility and uterine cancer, so they can be treated. In a conventional scenario, PCOS can be majorly classified into two types [1]:

- **Hidden PCOS**: The primary reason behind hidden PCOS is due to thyroid disease and low levels of iodine in the diet. However, once diagnosed, hidden PCOS can be treated within three to four.
- **Inflammatory PCOS**: PCOS that occurs due to an unhealthy diet and poor lifestyle is termed inflammatory PCOS and might cause headaches due to low levels of Vitamin C.

The entire functioning of the reproductive system in a female body depends on the hormonal levels which are further needed and responsible for conception and ovulation. The reproductive hormones that are majorly needed to keep a balance of the internal organs [2]:

- **Progesterone**: It is a female hormone secreted by ovaries once the ovulation cycle gets completed and the secretion is regulated by the LH hormone.
- **Estrogen**: It is a female hormone secreted by ovaries before the ovulation cycle begins and the secretion of estrogen is regulated by the FSH hormone.

Progesterone and estrogen are hormones secreted within the ovaries, whereas LH and TSH are hormones secreted through the pituitary gland. In addition to this, women with PCOS might indicate symptoms of infertility, miscarriages, and high levels of blood pressure (Kiruthika et al.; 2018). However, ultrasound imaging or ultrasonography can detect PCOS in a female body. The process includes capturing various parts of the ovaries through different angles so that a detailed study of the cysts can be performed and monitored. Figure 1 differentiates a healthy ovary from that a polycysticovary.

1.1 Background

PCOS is considered a significant hormonal ailment amongst females that tends to create an imbalance of reproductive hormones and leads to infertility. Such a disorder is observed in the reproductive phase, wherein the follicles of a female ovary tend to develop multiple cysts that grow on the outer edge of the ovary. The cysts increase in size with time and may grow in multiple phases. The cysts are fluid and are further termed follicles (Deswal et al.; 2020). Due to the development and growth of such follicles, the ovaries cannot produce mature eggs and, therefore, cannot release them on time. The untimely

[1]https://www.indiraivf.com/blog/types-of-pcos
[2]https://www.creative-diagnostics.com/blog/index.php/estrogen-and-progesterone/

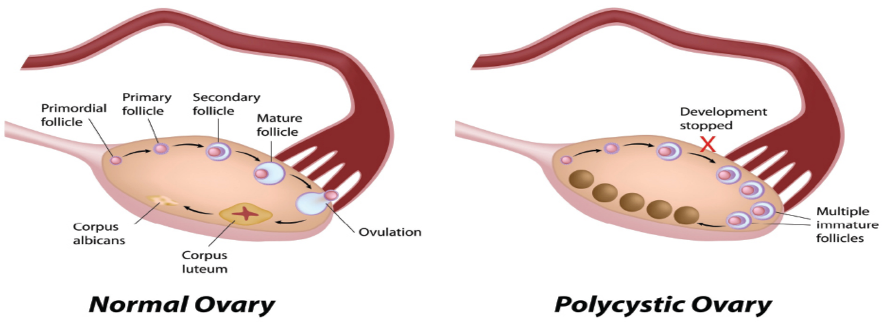


Figure 1: Visual Representation of a Healthy and a PCOS-Infected Ovary (Neuzil; 2014)

release of mature eggs impacts the menstrual cycle of a woman and would further lead to complications in conceiving. The symptoms of PCOS majorly include increased weight, growth of follicles on the ovary, excessive growth of body hair and facial hair, depression, mood swings and ovulation disorders such as irregular or heavy menstrual cycles. In addition, the presence of immature eggs leads to an increase in androgen levels of hormones in the body. It might further impact an individual's weight and expose the female to thyroid conditions. The androgen levels in a female body are responsible for regularising the menstrual cycle; hence an increase in its amount directly impacts the production and release of mature eggs. Early detection, weight control and adopting a healthy lifestyle are the only options to deal with the problem. On the other hand, detecting PCOS early might also save a woman from miscarriage. Hence, it can be said that PCOS is linked to multiple abdominal adiposities, including infertility in women, which might further lead to obesity, make her body resistant to insulin and increase cardiovascular risk factors. Therefore, it can be said that a woman might be a PCOS patient if she depicts two of the following three symptoms:

- Ovulation problems and irregularities
- Increased levels of androgen hormones
- A rise in the number of follicles on the outer edge of the ovary

As per a survey conducted in 2018, 2 in 5 women are exposed to PCOS, with 70 % of the Asian women population yet unable to avail the diagnostic cure to the disease. In another survey conducted by (Bharathi et al.; 2017), the author mentions that urban women are more likely prone to adapt to the disorder in comparison to rural women. The primary reason for this ratio is the lifestyle and associated stress found in urban women more than women in rural areas. Hence, a lifestyle change contributes as a significant factor in eliminating PCOS. According to another statistic from WHO, conducted in 2012, this disease has already victimized more than 120 million women, and the numbers are more likely to increase in the coming years (Ramamoorthy et al.; 2020).However, the detection of PCOS can be done using ultrasonography which uses the frequency of 2-15 MHz of sound waves and tends to generate images captured from the female ovary and other internal organs of the uterus. The process of conducting sonography is straightforward, painless, and inexpensive. A grey image is generated from ultrasound and consists of two ovaries that depict the presence of multiple cysts and follicles on the ovary. In the process of ultrasound sonography, the number and size of cysts can be depicted. A typical patient with mild PCOS may have 3- 4 cysts, whereas a severe PCOS patient may encounter more than 10 cysts. An indication of more than ten cysts is an alarming number, and the female must be immediately put to further treatment for the disorder to avoid major reproductive complications (Hassan and Mirza; 2020).

1.2 Motivation

To date, many treatments and approaches have been used to track the pattern of follicle enlargement and the formation of cysts in the female body. Despite extensive research, doctors and practitioners still lack a consensus that can be followed for every female patient. In addition, other factors such as high blood pressure levels, gynaecological disorders, uterine cancer, and cardiovascular diseases are much more likely to be caught by a female body exposed to PCOS. Apart from chronic diseases, the regular flow of a menstrual cycle is also highly affected due to PCOS. A patient might undergo irregular menses or no mature eggs released. This scenario might lead to problems of infertility and further possess a challenge for a woman to conceive. However, the only treatment for the disease is through a change in lifestyle and overcoming the issue of obesity. A change in lifestyle generally indicates the pattern of sleep, intake of healthy food, the amount of exercise done by the body followed by no intake of junk food, and complete avoidance of stress. It is worth noting that, until now, a specific treatment for the disease has not yet been detected. Hence, its detection at an early stage is mandatory to avoid further reproductive complications in the female body. Therefore, this serves as the driving factor of the presented thesis. The primary contributions of the research thesis are (i) pre-processing the dataset through feature selection so that only salient features of the ovary image are selected and processed further, (ii) building architecture to propose the workflow of the same (iii) implementing CNN and transfer learning models to accomplish higher levels of accuracy.

1.3 Research Objectives

The research of the proposed thesis is based on the theory of computer vision using CNN and transfer learning algorithms. In addition, the automation of the PCOS detection process using the mentioned algorithms. As a result, the research objectives of the proposed work are as follows, depending on the flow of the thesis:

- To automate the process of PCOS detection.
- To implement the same using CNN and transfer learning models to achieve higher accuracy.
- To compare the proposed model with that of the existing one.
- To comprehend the methodology of the proposed work.
- To overcome existing challenges and expand the future scope.

1.4 Research Questions

Following are the research questions narrated to implement the proposed thesis:

- How can the proposed model serve the purpose of benefitting the healthcare industry?
- What parameters would be chosen for the evaluation of accuracy?
- Which transfer learning models would be used?
- How can the scope of the proposed thesis be defined?

2 Related Work

The topic of the domain for the detection of the occurrence of PCOS in women has been widely researched by multiple scholars. Various techniques and methodologies have been implemented and tested to diagnose the disease early so that further complications and issues related to medical health can be pondered upon. PCOS is believed to be an endocrine ailment with multiple criteria involved in its diagnosis. In addition, an ultrasound gives detailed information on the number of follicles and cysts present, along with their respective size. This calculation of follicle distribution is necessary to examine so that the diagnostic criteria can be ruled upon and the primary reason for its occurrence can be detected. The process of conducting an ultrasound includes tracing the ovary manually through the machine and further counting the presence of cysts so that an automated process for its detection can be executed (Lawrence et al.; 2007). Implementing the respective algorithm helped recover follicle segmentation that was retrieved from ultrasound images. This segmentation included studying the follicles in a detailed manner through their features and further segmenting them using stereology. The obtained follicles were later stored as feature vectors, and in the last step PCOS classification was performed. The classification steps included categorizing the data as PCOS present and PCOS absent. The implemented method helped to automate the entire system, and 3 machine learning classifying algorithms were used. The process of follicle development in an ovary differs in each woman. It is through the development stage that the intensity of the disease is diagnosed. The stage of follicular development also tends to distinguish various forms of PCOS ailments. Hence, the diagnosis of such follicles is made through medical expertise that manually calculates the number of cysts and the follicles involved. This calculation and segmentation of the same tend to monitor further complications in reproducibility. However, to further control such complications, (Deng et al.; 2008) proposed work in the same domain to automate the system so that early disease detection could be performed. The ultrasound images were used as input and feature extraction and filtering using morphological concepts were implemented. The obtained output was labeled through a watershed model, and the respective target was selected for feature engineering. A clustering method was applied to data after feature engineering to identify follicular cysts. The method generated high levels of accuracy in detecting the same and thereby automated the entire implementation process to achieve an accuracy of 84%. However, because of its great relevance to PCOS, the proposed technique may not be applicable towards other multitarget identification issues.

(Vigil et al.; 2009), identified the presence of two cervical mucus, namely; gestagenic and oestrogenic. Such classification helps to understand women's crystallization characteristics, which in turn helps to identify cervical mucus at early stages. It would further be detected using ultrasound images obtained from women who already suffered from PCOS. Ten samples from a repository are used, of which four belong to women suffering from cervical mucus, and the rest suffered from regular PCOS. Due to the presence of crystallization in cervical mucus, the levels of progesterone and oestradiol tend to increase to higher levels of hormonal shoot-up. Hence, a differentiation factor was created to identify women suffering from cervical mucus and PCOS. One of the features that created a well-established differentiating factor was the suffering of an anovulatory cycle in women with PCOS. The differentiation in level was due to the diameter of the cervical mucus found in women. A circular diameter was observed in women suffering from cervical mucus, whereas a hexagonal diameter was observed in women suffering from PCOS.

2.1 Traditional Methods

Through the conduction of the literature survey, it was observed that PCOS was one of the significant reasons that add to the complications in anovulation and infertility. However, the inclusive criteria depicted that specific parameters were to be found by the experts and doctors through manual examination of the patient. Such examinations included the process of evaluation of various parameters that were to be previously described before the execution of the detection would occur. (Mehrotra, Chatterjee, Chakraborty, Ghoshdastidar and Ghoshdastidar; 2011), contributed his work in this domain and proposed an automated system that could diagnose the same in the early stages so that further complications could be avoided. The suggested model involves the implementation through the gathering process of ultrasound images of the ovaries and further conducting feature extracting techniques on the same do that all the respective features could be quickly evaluated. Once the statistical feature extraction was done, the classification process was done to segregate the models as PCOS positive or PCOS negative. For this purpose, the author implemented machine learning algorithms such as Logistic Regression and Naive Bayes. The automated process of detection was carried out on 450 samples of positive patients and 328 samples of adverse patients. It was observed that the model with the implementation of Logistic Regression generated higher levels of accuracy and achieved a total precision of 93%.

Through a survey, it was observed that almost 10 % of the women were affected by PCOS. It leads to high complications in their stage of infertility. (Saito and Ohmori; 2011) suggested an automated system to detect the same built on the conceptual theory of mathematical grounds that could detect the presence of sterilities in a woman. The method was solely entitled to initiate the same treatment so that PCOS detection could be done at the early stages and diagnosed for proper treatment in women. The performance of the model was based entirely on the functioning of machine learning algorithms and models that helped diagnose the same among PCOS patients. The model tends to illustrate the levels of LH hormone in the female body and further suggests a specific treatment. Since the LH hormone regulates a woman's menstrual cycle, its calculation and monitoring were essential. Therefore, to monitor such activities and to create a balance of hormonal levels, certain medications are required and thereby prescribed by the doctor. The doctor's dosage helps maintain the LH levels in a female body.

In another work suggested by (Mehrotra, Chakraborty, Ghoshdastidar, Ghoshdastidar and Ghoshdastidar; 2011), he proposed automating the system using the pathognomonic behaviour of the follicles and the cysts in the female body. The behaviour was used to derive respective patterns and determine follicular arrangements in the ovary. However, the entire implementation process was carried out using the dataset from GDIFR University in Kolkata. The samples were collected and tested upon women belonging to the age group of 25-35 years of female patients. In the later stages, ultrasound sonography was also conducted using a 7MHz transducer to diagnose the same. The final verification of the reports and the output data so generated was performed by clinical experts and done by a gynaecologist. The images obtained were initially pre-processed, and respective machine learning-based methodologies were used. A morphological image of the same was observed, and contrast enhancement was done using the multiscale approach.

The occurrence of the disease tends to affect the reproductive health of a female body, which can either be treated through medication or surgery. However, a manual examination of the female body needs to be done by a clinician. (Rihana et al.; 2013) suggested a machine learning system competent in automating the process of determining cyst intensity. For implementation, 25 image samples were obtained from the repository containing the ovaries' images and the conversion to grey-scale images was performed. This conversion enhanced the entire process of image binarization and created a contrast in the images so obtained. In the next stage, the morphology process was done, and the extracted features were later used to characterize the size and number of follicles obtained. The author implemented the proposed algorithm using SVM as the classifier and later validated the entire process using the evaluation parameters such as accuracy and precision factors. A total accuracy of 91 % was obtained. The evaluation would have been much more efficient had there been a larger number of images trained.

Another research (Sitheswaran and Malarkhodi; 2014), proposed a work in the same domain and suggested diagnosing the development of follicles in the ovary using manual examination followed by imaging of ultrasound sonography. In order to identify and diagnose the increasing number of follicles, a mathematical algorithm named the object growing algorithm was implemented. The entire process of implementation was undergone through two stages that initially involved pre-processing the data followed by identifying the size of the cysts. The noise and irrelevant data present in the dataset were removed and discarded so that accurate classification of the disease could be done. In the later stages, a median filter was also adopted to reduce redundancy. Once the feature extraction process is completed, a watershed algorithm is used so that the additional labelling process can be performed; this algorithm is responsible for extracting the size and number of follicles so that local minima can be calculated. The final implementation of the 'object growing' algorithm is done to detect and identify the presence of cysts in the female body so that further disease prevention can be done. Since the recognition of a high number of follicles can lead to ovarian failure, its detection is mandatory so that infertility can be treated. As mentioned earlier, its detection can be done using ultrasound, and meaningful information can be retrieved through sonography.

The work of (Kumar and Srinivasan; 2014) was based on revision performed through the Vase Method so that a detailed version of follicle segmentation and detection could be done. For implementation, 55 sample images of the ovary were collected from women between 25 and 35. These women suffered from major PCOS. The author proposed the implementation of his work through the usage of machine learning algorithms. Three algorithms were used, and the model implemented using SVM generated the highest accuracy results.

2.2 Machine Learning Methods

(Setiawati et al.; 2015) proposed the implementation of automatically diagnosing the disease using the clustering method through images. The method was responsible for performing the segmentation of follicles using an optimization method. A non-parametric function was developed through the optimization process to derive RMSE values. The values attained were used to create an index value so that the final calculation of accuracy could be done. The final results were used to match the images obtained through ultrasound sonography, and precision values were calculated using the same. In the next stage, the adopted optimization method was enhanced through particle swarm optimization (PSO), and the performance of follicle size and its growing number were monitored.

According to one of the surveys conducted through the National Institutes of Health (NIH), it was observed that almost 8-12 % of women suffered from PCOS disease, which was not recognized at the right stage. (Purnama et al.; 2015) suggested implementing a model based on identifying the follicle size and determining its impact on a female body. It was observed that the number of follicles and the larger the size of the follicle, the more difficult it is for a female body to conceive. Thereby resulting in reproductive issues in the female body. Hence, the author proposed to diagnose the issue related to the growing size of follicles in the ovary by conducting experiments using machine learning algorithms. In the initial stage, the technique of data pre-processing is done so that irrelevant and redundant data can be removed. The pre-processing stage also involved the data visualization stage. An equalization histogram was created to depict ovarian diseases. In the next stage, morphological changes and filters were passed through the images obtained from ultrasound sonography. Later, segmentation was performed on the obtained results from the pre-processing stage. This phase involved the process of cropping and edge detection. Once the follicle images were labelled through the segmentation method, the Gabor wavelet technique was implemented to extract the features. For implementation, two datasets were used, including 40 images and 26 images, respectively. The first dataset involved women suffering from PCOS, whereas the second dataset involved women suffering from cervical mucus. The entire implementation was carried out on three machine learning algorithms, and the model with SVM generated higher accuracy results.

In research proposed by (Chen et al.; 2019), a manual examination of the female body was initially done by an expert or a doctor. In the later stages, ultrasound sonography was conducted on a dataset of 242 women who suffered from the disease of PCOS. A cardiovascular examination of the entire female body was done to verify blood pressure levels. A radial pulse spectroscopy was also conducted for women with high cysts in their ovaries. Certain women, who had surpassed their age of childbearing phase, also gave their samples and their images were also equally examined. A pre-processing data phase was conducted, and Fourier transformation was used to fulfil the feature extraction phase. A total of 5 machine learning algorithms were used, and the algorithm with the implementation of KNN generated higher levels of accuracy. The proposed research work also mentioned that women with high body indexes illustrated high blood pressure (C2 and C4) and vice-versa.

Studies and multiple research works have shown that PCOS has been witnessed to be an endocrinopathy disorder. For that purpose, it has been widely researched. Detecting PCOS was implemented using two machine learning models by (Xie et al.; 2020) that were characterized by random forest and decision trees. In addition to machine learning models, the author also imposed the concepts of deep learning models in the same research work. An Artificial Neural Network (ANN) was used as the deep learning model, and gene biomarkers were highlighted to diagnose the model's functioning. In the initial stage, a dataset was collected from a respective repository that contained 58 regular and 75 PCOS ovarian images of female patients. The gene biomarkers helped to assist gene expression so that the labelling of the dataset was done efficiently. In addition to this, five more sets of databases were used so that screening of female patients could be done using gene expression. The five datasets used in the implementation also was divided into train, test and validation data. Fifteen gene keys were established and used to diagnose the disease using random forest and decision trees, whereas 12 gene keys were used to diagnose the disease using ANN. The dataset is classified, and the samples were divided and bifurcated as PCOS-positive and PCOS-negative patients. The validation dataset was implemented to calculate the weight of the neurons involved in the diagnosis through the ANN procedure.

The implementation diagnosis of the same was proposed by (Khan Inan et al.; 2021) using a gradient-boosting algorithm. The significant symptoms of PCOS are high blood pressure levels and Type 2 diabetes. Hence, the proposed solution was implemented to detect and diagnose the disease using gradient boosting algorithms such as XGBoost and Gradient Boosting. The dataset of PCOS patients depicting the growing number of follicle sizes was initially collected from the Kaggle repository, and all the classifiers, including machine learning and boosting algorithms, were trained on the model. In the later stages, the boosting algorithms were further accompanied by algorithms such as K-Nearest Neighbors (KNN), Random Forest (RF), Support Vector Machine (SVM) and Naive Bayes (NB). The performance implementation of the model was later accompanied by the Chi-Square test followed by SMOTE, which was used to evaluate the parameters for accuracy and precision. In the later stages, cross-validation with ten folds was implemented with the desired accuracy of 98 %.

The detection process of PCOS has been widely supported through ultrasound sonography. The manual examination has helped to treat multiple patients who have been suffering from infertility and ovarian disorders. The images scanned from ultrasound have helped to gather information on the size and the number of follicles that have been present in the ovary. It was well observed that the primary reason for an imbalance of hormones in a female body is the growing number of cysts in the ovary. (Madhumitha et al.; 2021) proposed the implementation process of PCOS detection through image segmentation that involved diagnosing the Region of Interest (ROI). This region was initially separated from the images in the background, enhancing an image's posture. In the later stages, the author implemented the same using four algorithms: KNN, SVM, NB and Logistic Regression. All the algorithms were implemented, and trained individuals and the model implemented using NB generated the highest accuracy result of 98 %.

2.3 Observational Drawbacks

The following are the drawbacks of the existing systems so witnessed during the conduction of the literature survey:

- Studies that included the detection of PCOS using traditional machine learning methods did not include the psychological parameters of women.

Table 1: Literature Review Summary

Author	Limitations	Algorithm	Accuracy
(Ahmetašević et al.; 2022)	Class Imbalance was present and the model was trained without balancing the dataset.	Artifical Neural Network (ANN)	96.1%
(Bharati et al.; 2020)	Most crucial variables were dropped from the model building during feature selection.	Random Forest (RF) + Logistic Regression	91.01%
(Cahyono et al.; 2017)	Dropouts were applied in all layers to increase the accuracy during model building.	CNN	76.36%
(Deng et al.; 2008)	The model is limited to the number of targets. Does not suit multiple target detection problems.	Boundary Vector Field (BVF)	84%
(Denny et al.; 2019)	Principal Component Analysis was performed. Difficult to select the biological parameters.	Random Forest Classifier	89.02%
(Deshpande and Wakankar; 2014)	A Wiener filter is used which is difficult to restore the random nature, requiring more time.	Support Vector Machine (SVM)	95%
(Hijab et al.; 2019)	Transfer learning models were fine-tuned to detect Breast tumours.	VGG 16 CNN	97%
(Khan Inan et al.; 2021)	Outliers are removed, and the techniques are not mentioned.	Extreme Gradient Boosting (XGBoost)	98%
(Madhumitha et al.; 2021)	The implemented approach provides better accuracy using machine learning algorithms.	Naive Bayes	98%
(Purnama et al.; 2015)	Dataset is too small and is imbalanced.	Radial Basis Function (RBF)	82.55%
(Xie et al.; 2020)	Gene Expression Dataset limited to sample tissue	Random Forest. + ANN	72%

- Integration between the detection of PCOS and women's mental health was absent.
- Existing techniques majorly involved the implementation of statistical tools (ML and DL-based algorithms); no observation of soft computing approaches such as fuzzy logic was witnessed.
- A real-time application that could detect PCOS in the early stages and connect the doctor with the patient is yet to be viewed.

Given the information shown above and in Table 1, the majority of PCOS research focuses on the disease's metabolic and clinical characteristics. Taking into consideration the drawbacks of the existing system; the proposed thesis tends to highlight the physical and psychological changes occurring in women. A technique of combining deep learning algorithms with transfer learning models is employed in this thesis so that the implementation of statistical tools along with soft computing approaches can be achieved.

3 Methodology

Knowledge Discovery in Database (KDD) refers to the process of discovering vast amounts of data using mining techniques so that many research projects in Deep Learning, Artificial Intelligence, and Machine Learning can be carried out. It has different stages like domain understanding, selection, pre-processing, transformation, data mining, evaluation and visualizing the insights produced as shown in Figure 2. Because of the increasing

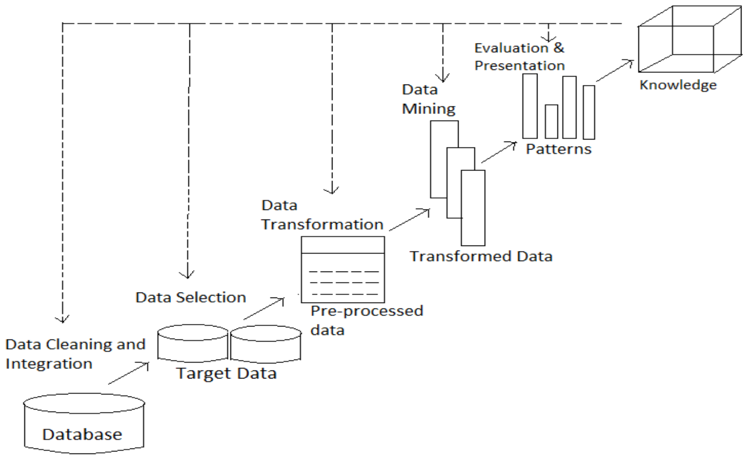


Figure 2: Steps Involved in KDD Process (TechBlogMU; 2018)

competition and demand for decision support, KDD has been utilized to categorize women as PCOS infected or not utilizing transfer learning techniques, and CNN gave the ultrasound image (Tan et al.; 2015).

4 Design Specification

Figure 3 depicts the architectural diagram of the thesis and its different stages.

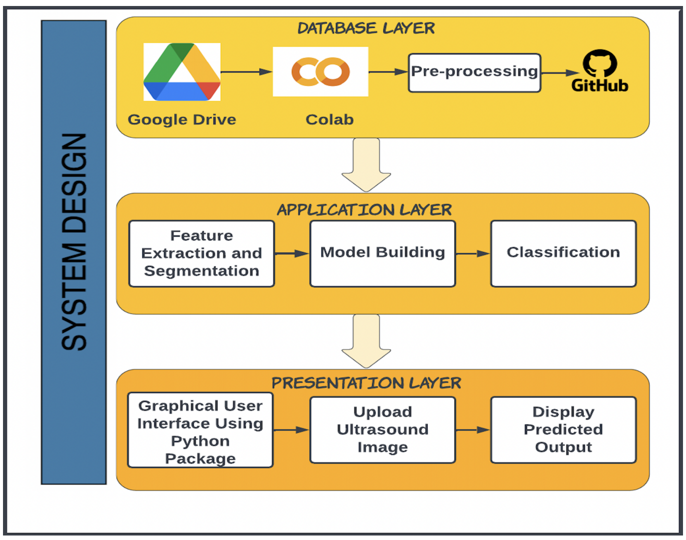


Figure 3: Architectural diagram of the proposed system

1. Database Layer: This stage includes the collection of the dataset on google drive that would contain PCOS positive and negative images of a female patient. Since the images are gathered from different sources, they are resized to a respective resolution so that the model can predict whether PCOS is present. The implementation of the data then occurs on google colab, followed by the data pre-processing phase. The progress in code is saved in the gitlab.

2. Application Layer: The next stage involves a training phase, wherein the model undergoes the feature engineering process of image segmentation and extraction. Once this is completed, a model is built and classified for further prediction

3. Presentation Layer: This forms the last stage of implementation that includes a GUI window; wherein an ultrasound image would be uploaded, and the model so deployed would be able to predict the output and display it as PCOS positive or PCOS negative.

5 Implementation

The process of detecting a female patient with PCOS is a time-consuming and tedious task. Therefore, the presented research work proposes automating the implementation of diagnosing the disease so that it can further be assisted through manual examination by medical practitioners and experts. This would help to identify the disease at the right stage and prevent its occurrence in the later stages so that ovarian diseases can be cured. The implemented method tends to predict the disease's occurrence through a

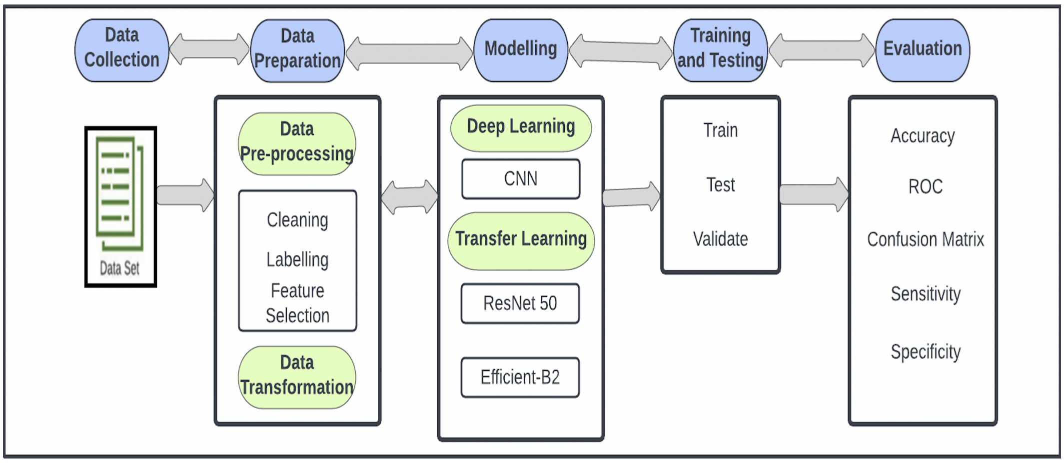


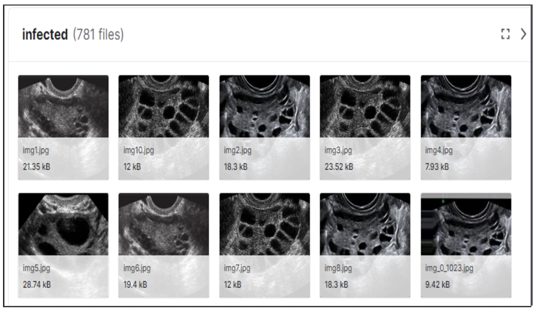
Figure 4: Methodological workflow of the proposed system

dataset provided to the system. Several research scholars have employed the conceptual framework of deep learning and machine learning algorithms to apply the workings of the same throughout the literature review. However, this research puts forward the diagnosis of PCOS amongst women with ovarian diseases using deep learning and models of transfer learning. CNN is the deep learning algorithm, followed by ResNet-50 and EfficientNet-B2 as the pre-trained transfer learning-based model. Since the dataset used for implementation is limited, the thesis proposes the architectural implementation of data augmentation to generate more images to obtain higher accuracy and precision.

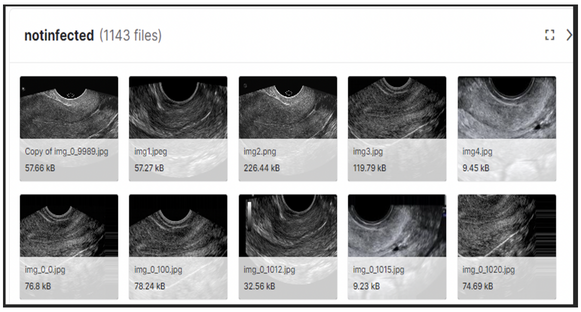
Figure 4 depicts the methodological workflow of the proposed implementation. The process flow has five steps: data collection, preparation, model building, training and testing, and evaluation. Different algorithms from deep learning and transfer learning are implemented in the model-building stage. Different pre-processing techniques and data transformation are performed in the data preparation stage. Furthermore, evaluation is implemented using accuracy, ROC and other metrics.

5.1 Data Collection

The entire implementation of the research study began with collecting the dataset from the Kaggle repository. The repository consisted of train and test set files. In addition to the train and test files, the dataset consisted of 781 ultrasound images of PCOS-infected women and 1143 images of PCOS not infected women. The entire dataset consisted of images in the form of ultrasound scans performed on women. The images of PCOS-infected women displayed a growing number of follicles, leading to the detailed examination of the females done through the manual process. On the other hand, the images obtained from PCOS not infected women had normal ovaries with regular shapes. Figure 5a and Figure 5b illustrate the obtained dataset with PCOS-infected patients.



(a) PCOS infected patients



(b) PCOS not infected patients

Figure 5: Types of Data

5.2 Data Preparation

This stage of implementation verifies the repository for any NULL or missing values from the dataset and tends to discard redundant columns so that the overall efficiency of the system program can be elevated. The process, however, includes the following stages of implementation:

1. **Data Labelling**: Data labelling includes allocating tags to their respective attributes depending on their subsets of data. The labelling process occurs so that PCOS infected and not infected can further be differentiated after getting converted to their respective binary representations. The labelling process also makes it easier for deep learning and transfer learning algorithms to execute their methodology on the dataset, thereby increasing the system performance of the model. However, this process converts non-categorical features to their respective numerical values.

2. **Data Binarization**: The data binarization process includes converting selected attribute labels to their respective binary formats, which are further represented as

0 or 1. It is a mandatory process wherein the obtained images from the repository are represented through 0 for PCOS negative (not infected) and 1 for PCOS positive (infected). This process is done so that the system can comprehend the conversion of the image to binary form. However, this process includes using categorical functions to convert attributes to binary form.

5.3 Data Augmentation

The data augmentation process includes replicating, creating and developing multiple images from a single image. It is done by performing operations such as horizontal flipping, rotating, saturation, transforming, and zooming so that a single image can generate multiple images. The conceptual theory of data augmentation does not include generating new data but rather replicating the original data to produce multiple copies of the same. The data augmentation process enables the dataset's enlargement since the dataset used in the thesis implementation is limited. Larger the dataset, the higher the chances of accurate result generation.

5.4 Data Modelling

The architecture of the proposed model is built on the concepts of deep learning and transfer learning used to implement the model. In addition to this execution, augmentation is also performed on the dataset to increase the intensity and diversity obtained from the dataset. For implementation following are the algorithms that have been used:

- **CNN**: CNN is the most commonly accepted and widely used deep learning algorithm used to generate high levels of accuracy that were not obtained through machine learning algorithms. The implementation of a CNN is based on hidden layers that resemble the working process of a neuron. Multiple layers are usually associated with a CNN. However, it includes four major layers responsible for predicting the model's outcome. The initial layer begins with the convolutional layer, also referred to as the input layer and is further responsible for gathering input data from the dataset. Data obtained from this layer is convolved using mathematical operations and fed to the preceding layer, which forms the pooling layer. The pooling layer is majorly responsible for combining all the respective inputs, pooling them together, and feeding them to the next layer. The third layer so involved is termed the activation layer. It is one of the most important layers wherein the activation function, such as a Sigmoid or ReLu is used depending on the weight of the associated neuron. The final layer is termed the fully connected layer. It tends to combine the associated biases and weights concerning the neurons so that forward and backward propagations can occur in accordance with the weights assigned.

- **ResNet-50**: One of the commonly used transfer learning-based CNN models is ResNet50. It is a convolutional-based model that consists of 50 layers stacked on top of one another. An important feature of ResNet50 is that it overcomes the vanishing gradient problem. The architecture also consists of short connections termed "skip" that tend to overcome some part of execution and iterates the model with essential and required steps of execution.

- **EfficientNet-B2**: EfficientNet-B2 is an extended and enhanced version of the convolutional neural network (CNN) that falls under the deep learning architecture.

The model uses a scaling method that requires uniformity in all scales concerning its dimensional factors, such as depth and width. In addition to dimensional factors, they utilize a compound coefficient responsible for managing the resolution of the image given as the input. Unlike the working concept of a traditional CNN that uses scaling factors, the implementation of EfficientNet-B2 depends on scaling coefficients so that the final resolution of the image is not distorted. For instance, if the computational resource to be used is magnified to 2N times, the overall depth of the network increases by N, whereas the width increases by βN, where α and β are considered to be the respective coefficients that would eventually be magnified from the original model. EfficientNet-B2 uses a compound coefficient θ to uniformly scale network width, depth, and resolution in a principled way.

5.5 Data Visualization

The approach of depicting the flow of data is known as data visualization. Graphs, plots, bar charts, descriptive statistics, and images are widely used in their representation because they help extract information from the input. This generated knowledge aids in comprehending a data incident that takes place. The data visualization process begins once the data has been extracted, processed, and cleaned. A built-in package known as count-plot is used to count different mathematical operations on the dataset. Figure 6 depicts the number of PCOS-infected and PCOS not infected images obtained from the dataset.

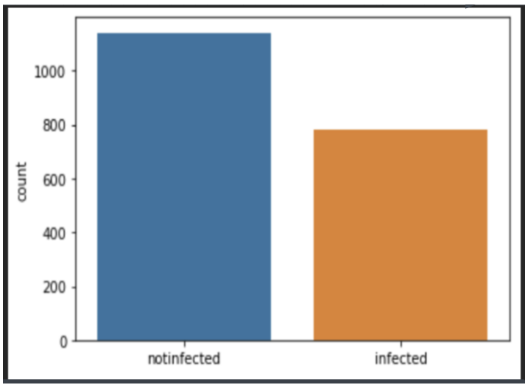


Figure 6: Data Visualization of PCOS class

5.6 Model Training

The obtained dataset undergoes the process of implementation using three algorithms, as mentioned above. Once the process is completed, the data is further trained, tested and validated so that final accuracies of the system model can be generated. The initial dataset consisted of train and test files with 1922 ultrasound images available for training purposes and 196 images available for testing purposes. In addition to this, the dataset comprised 781 infected images of PCOS patients and 1143 images of not-infected patients. The entire dataset was split into 90% data for training purposes and 10% data for testing purposes. Next, the data was also validated to elevate the overall performance of the system model. For validation purpose, 10% of data is taken from the training folder.

Three different algorithms such as ResNet-50, EfficientNet-B2 and CNN was implemented

Table 2: Fine Tune CNN Layers

Parameters	Value
Optimizer	Adam
Activation Function	ReLu
Filter Size Used for Feature Extraction	15,12,8

to detect the PCOS. For CNN, few parameters are fine-tuned to make the model efficient as shown in Table 2. Different Padding and kernel size parameter gives the best accuracy. Different activation function for feature extraction is implemented, and relu is chosen as the effective one because of better accuracy (compared to sigmoid and softmax). Max pooling parameter with different pooling sizes is implemented to achieve better accuracy than average pooling. In contrast to the SGD optimizer, the Adam optimizer achieves better accuracy.

5.7 Evaluation Metrics

The conduction of the evaluation matrix is a mandatory task and is performed so that evaluation parameters can be defined for better results. The matrix and the parameters used for evaluation represent the system's overall performance so that the efficiency can be estimated and improved. However, the parameters of each system model need to be defined following the methodological workflow of the system. For the implementation of the proposed thesis, the following is the list of evaluation parameters:

1. **Confusion Matrix**: The performance of any system model is calculated using the values obtained from the confusion matrix. It is usually a representation of tabular format wherein specific values are filled with respect to the actual outcomes and the outcomes so predicted.

2. **Classification Report**: classification report provides details on the values obtained from accuracy 1, recall 4, F1Score 3 and precision 2 factors. The table below represents the formulae for calculating the respective parameters:

$$Accuracy = \frac{TN + TP}{TP + FP + TN + FN} \tag{1}$$

$$Precision = \frac{TP}{FP + TP} \tag{2}$$

$$F_1Score = 2\frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{3}$$

$$Recall = \frac{TP}{FN + TP} \tag{4}$$

3. **Specificity**: specificity is a ratio that gives information of negative values so obtained with respect to all the occurrence of positive instances in the fraction so present.

4. **Sensitivity**: sensitivity is a ratio that gives information of positive values so obtained with respect to all the occurrences of negative instances in the fraction so present.

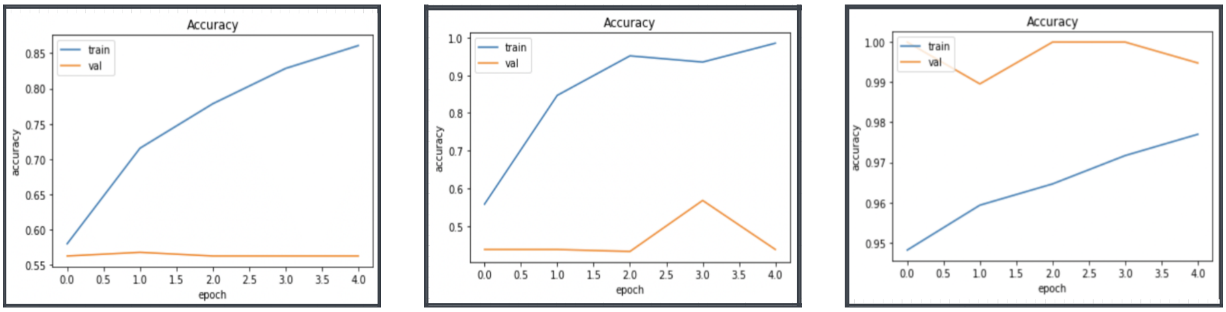
6 Experimental Analysis and Results

This section of the thesis includes the experimental analysis performed on the system model to obtain the desired accuracy using certain evaluation parameters, as stated above.

6.1 Algorithms depicting Accuracy Vs Loss Graph

A graph that illustrates the accuracy generated with respect to the error loss is termed the accuracy VS loss graph. The model depicting the highest accuracy with minimal error loss is declared an optimized model. Figure 7 and Figure 8 indicates the respective accuracy and loss values so obtained through the implementation of different algorithms.

The graphs have been generated using Adam as the optimizer for five epochs each. It depicts an increase in validation accuracy and an observed decrease in error loss. However, the dataset used included ultrasound images generated through the augmentation process.

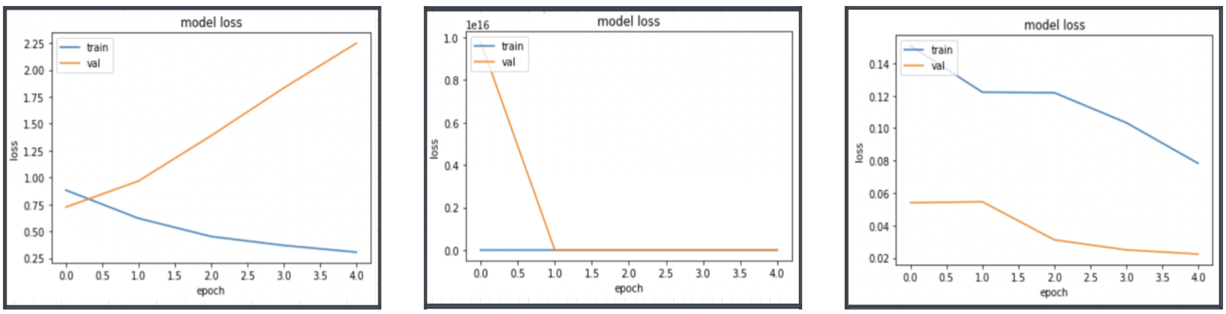


(a) RESNET (b) EfficientNet-B2 (c) CNN

Figure 7: Train and Validation Accuracy Graph

6.2 Algorithms depicting Confusion Matrix

A confusion matrix typically represents the values obtained after the system model's implementation. It compares values generated after prediction to that of the original values. Figure 9 indicates the confusion matrix of so obtained through the implementation of different algorithms. Figure 9a represents TP, FP, TN and FN values obtained for the ResNet-50. It depicts that 96 PCOS cases have been predicted to be positive and tend



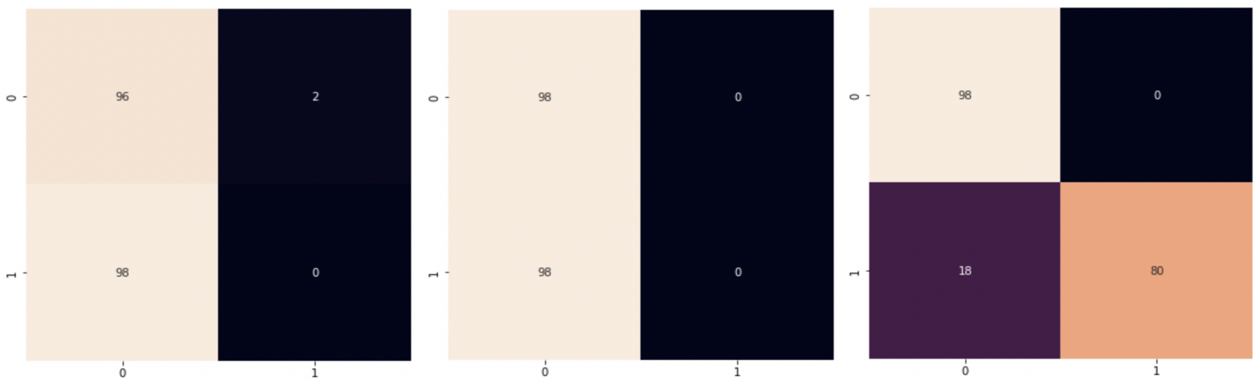
(a) RESNET (b) EfficientNet-B2 (c) CNN

Figure 8: Loss Graph of Train and Validation Data

to be positive in actual value. However, the number of false and true negative is higher than the positive cases.

The block representation of a confusion matrix generated through the implementation of EfficientNet-B2 has been illustrated in Figure 9b. The graph represents the values obtained for TP, FN, TN and FP. It depicts that 98 PCOS cases have been predicted to be negative and tend to be negative in actual value. It is also evident that the model predicts only half of the PCOS cases correctly.

The block representation of a confusion matrix generated through the implementation of CNN has been illustrated in Figure 9c. The graph represents the values obtained for TP, FN, TN and FP. It depicts that 80 PCOS cases have been predicted to be positive and tend to be positive in actual value. Similarly, 98 cases have been predicted to be non-infected and tends to be negative in actual value.



(a) RESNET (b) EfficientNet-B2 (c) CNN

Figure 9: Confusion Matrix

6.3 Classification Reports

From the classification table generated in Figure 10a, it can be observed that 49 % of PCOS not infected female patients have been diagnosed, and 0 % of PCOS infected have been detected. The same values can be deduced for recall and F1 score factors. In addition to the values so obtained, it can be observed that the model implemented using ResNet50 tends to generate an overall accuracy of 49 %.

From the classification table generated in Figure 10b, it can be observed that 50 % of PCOS not infected female patients have been diagnosed, and 0 % of PCOS infected have been detected. It is also evident from the precision, recall and f1-score values that the model is biased towards the non-infected cases. In addition to the values so obtained, it can be observed that the model implemented using EfficientNet-B2 tends to generate an overall accuracy of 50 %.

From the classification table generated in Figure 10c, it can be observed that 84 % of PCOS not infected female patients have been diagnosed, and 100 % of PCOS infected have been detected. The same values can be deduced for recall and F1 score factors. In addition to the values obtained, it can be observed that the model implemented using CNN generates an overall accuracy of 91 %. Through the implementation of the classification report, it can be concluded that; to fulfil the purpose of PCOS detection, the system model of CNN has generated the highest accuracy of 91 % compared to the accuracy generated by ResNet-50 and EfficientNet-B2 to be 52 % and 50 % respectively.

	precision	recall	f1-score	support
0	0.49	0.98	0.66	98
1	0.00	0.00	0.00	98
accuracy			0.49	196
macro avg	0.25	0.49	0.33	196
weighted avg	0.25	0.49	0.33	196

(a) RESNET

	precision	recall	f1-score	support
0	0.50	1.00	0.67	98
1	0.00	0.00	0.00	98
accuracy			0.50	196
macro avg	0.25	0.50	0.33	196
weighted avg	0.25	0.50	0.33	196

(b) EfficientNet-B2

	precision	recall	f1-score	support
0	0.84	1.00	0.92	98
1	1.00	0.82	0.90	98
accuracy			0.91	196
macro avg	0.92	0.91	0.91	196
weighted avg	0.92	0.91	0.91	196

(c) CNN

Figure 10: Classification Reports

6.4 Algorithms depicting Sensitivity and Specificity

The values of sensitivity and specificity so obtained from different algorithms depict their ratio of positive values to negative instances and negative values to positive instances.

The sensitivity and specificity of ResNet-50 in Figure 11a are 0 and 0.979, respectively, for PCOS-negative patients. The values obtained indicate that the ratio of positive values to negative instances for PCOS negative instances (sensitivity) is 0.0, whereas the ratio of negative values to positive instances (specificity) is observed to be 0.97 for PCOS negative instances.

The sensitivity and specificity of EfficientNet-B2 in Figure 11b are observed to be 0.00 and 1.0, respectively, for PCOS-negative patients. The values obtained indicate that the ratio of positive values to negative instances for PCOS negative instances (sensitivity) is observed to be 0.00, whereas the ratio of negative values to positive instances (specificity) is observed to be 1.0 for PCOS negative instances. The values so obtained indicate that the ratio of positive values to negative instances for PCOS negative instances (sensitivity) is observed to be 0.816, whereas the ratio of negative values to positive instances (specificity) is observed to be 1.00 for PCOS negative instances. The sensitivity and specificity of CNN in Figure 11c are observed to be 0.816 and 1.00, respectively, for PCOS-negative patients. Therefore, the implementation of CNN has generated 91 % accuracy with minimal error loss.

	class	sensitivity	specificity
0	0	0.000000	0.979592
1	1	0.979592	0.000000

(a) RESNET

	class	sensitivity	specificity
0	0	0.0	1.0
1	1	1.0	0.0

(b) EfficientNet-B2

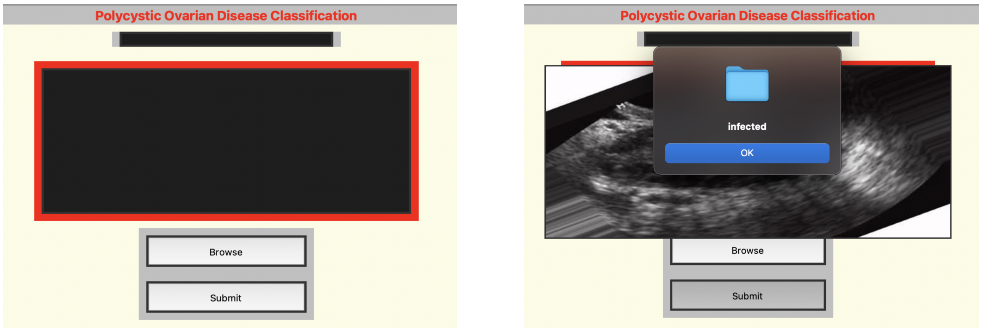
	class	sensitivity	specificity
0	0	0.816327	1.000000
1	1	1.000000	0.816327

(c) CNN

Figure 11: Sensitivity and Specificity Reports

6.5 Graphical User Interface (GUI)

Since CNN is suited to be the best and most optimized model, it is further used in the GUI so that an interface can be created to accomplish fine-tuned accuracy. Python's Tkinter and pillow packages have been adopted to implement the GUI with no service delay. The snippets below depict the interactions done on the GUI. Figure 12a illustrates the GUI dashboard to be used as an interface to predict the given file as PCOS positive or PCOS negative. For this purpose, the file is uploaded using the browse button and further submitted for prediction. The ultrasound image is uploaded using the browse button and is submitted for prediction to be made. The GUI predicts that the uploaded file is infected, thereby targeting that the patient is PCOS positive as shown in Figure 12b.



(a) Dashboard of the GUI (b) Prediction of the Uploaded File

Figure 12: Testing the Model Using GUI

6.6 Discussion

The research aims to evaluate the effectiveness of a newly proposed CNN model's effectiveness based on existing health sector findings cited in the literature review. Based on the evaluation results, the CNN model outperforms the other transfer learning methods since they fail to anticipate false positive scenarios. Furthermore, the accuracy of transfer learning models must be improved. According to Figure 10, CNN detects PCOS instances far better, with a mistake of only about 15 cases and an accuracy of roughly 91%. It might be regarded as an advancement in the detection of PCOS. Because diagnosing PCOS at an early stage allows doctors to begin treatment and cure the problem.

7 Conclusion and Future Work

The diagnosis and classification of the medical condition known as PCOS is a widely explored topic. However, it is vital to recognize an increasing number of cysts in the ovary so that subsequent ovarian disorders are avoided, and women do not experience infertility. The dataset used in this thesis was obtained from the Kaggle repository and included train and test files for 781 PCOS-infected patients and 1143 PCOS-uninfected individuals. Because the entire dataset was deemed limited, the data augmentation process was carried out to expand the overall data. Furthermore, a novel strategy for diagnosing the same is given, using deep learning algorithms such as CNN, ResNet-50, and EfficientNet-B2. A confusion matrix, accuracy graph, classification report, and sensitivity factors were among the evaluation metrics chosen to improve the system's overall performance. ResNet-50 produced 49% accuracy, followed by EfficientNet-B2, which produced 50% accuracy, and CNN, which produced the last and maximum accuracy of 91% . As a result, CNN was selected as the best model for analyzing the presented thesis.

However, future studies will entail collecting much larger data to attain higher accuracy findings before implementing the augmentation technique. Furthermore, more complex algorithms with more hidden layers can be utilized to detect PCOS. In addition, authentication methods can be implemented for the application, and the same can be deployed to the cloud environment.

References

Ahmetašević, A., Aličelebić, L., Bajrić, B., Bečić, E., Smajović, A. and Deumić, A. (2022). Using artificial neural network in diagnosis of polycystic ovary syndrome, *2022 11th Mediterranean Conference on Embedded Computing (MECO)*, pp. 1–4.

Bharathi, R., Swetha, S., Neerajaa, J., Madhavica, J., Moorthy, J., Santhanam, R., S., R. and B., U. (2017). An epidemiological survey: Effect of predisposing factors for pcos in indian urban and rural population, *Middle East Fertility Society Journal* **22**.

Bharati, S., Podder, P. and Hossain Mondal, M. R. (2020). Diagnosis of polycystic ovary syndrome using machine learning algorithms, *2020 IEEE Region 10 Symposium (TENSYMP)*, pp. 1486–1489.

Cahyono, B., Adiwijaya, Mubarok, M. S. and Wisesty, U. (2017). An implementation of convolutional neural network on pcos classification based on ultrasound image, *2017 5th International Conference on Information and Communication Technology (ICoIC7)*, pp. 1–4.

Chen, C.-Y., Chang, C.-W., Wang, S.-H. and Wang, G.-C. (2019). The effects of polycystic ovary syndrome on cardiovascular system in women of childbearing age, *BIBE 2019; The Third International Conference on Biological Information and Biomedical Engineering*, pp. 1–4.

Deng, Y., Wang, Y. and Chen, P. (2008). Automated detection of polycystic ovary syndrome from ultrasound images, *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* pp. 4772–4775.

Denny, A., Raj, A., Ashok, A., Ram, C. M. and George, R. (2019). i-hope: Detection and prediction system for polycystic ovary syndrome (pcos) using machine learning techniques, *TENCON 2019 - 2019 IEEE Region 10 Conference (TENCON)*, pp. 673–678.

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

Deswal, R., Narwal, V., Suneja, A. and Pundir, C. (2020). The prevalence of polycystic ovary syndrome: A brief systematic review, *Journal of Human Reproductive Sciences* **13**: 261.

Hassan, M. and Mirza, T. (2020). Comparative analysis of machine learning algorithms in diagnosis of polycystic ovarian syndrome,, *International Journal of Computer Applications* **975**.

Hijab, A., Rushdi, M. A., Gomaa, M. M. and Eldeib, A. (2019). Breast cancer classification in ultrasound images using transfer learning, *2019 Fifth International Conference on Advances in Biomedical Engineering (ICABME)*, pp. 1–4.

Khan Inan, M. S., Ulfath, R. E., Alam, F. I., Bappee, F. K. and Hasan, R. (2021). Improved sampling and feature selection to support extreme gradient boosting for pcos diagnosis, *2021 IEEE 11th Annual Computing and Communication Workshop and Conference (CCWC)*, pp. 1046–1050.

Kiruthika, V., Sathiya, S. and M.M., R. (2018). Automatic texture and intensity based ovarian classification, *Journal of Medical Engineering & Technology* **42**(8): 604–616. PMID: 30932717.

Kumar, H. P. and Srinivasan, S. (2014). Segmentation of polycystic ovary in ultrasound images, *Second International Conference on Current Trends In Engineering and Technology - ICCTET 2014* pp. 237–240.

Lawrence, M. J., Eramian, M. G., Pierson, R. A. and Neufeld, E. (2007). Computer assisted detection of polycystic ovary morphology in ultrasound images, *Fourth Canadian Conference on Computer and Robot Vision (CRV '07)* pp. 105–112.

Madhumitha, J., Kalaiyarasi, M. and Ram, S. (2021). Automated polycystic ovarian syndrome identification with follicle recognition, *2021 3rd International Conference on Signal Processing and Communication (ICPSC)* pp. 98–102.

Mehrotra, P., Chakraborty, C., Ghoshdastidar, B., Ghoshdastidar, S. and Ghoshdastidar, K. (2011). Automated ovarian follicle recognition for polycystic ovary syndrome, *2011 International Conference on Image Information Processing* pp. 1–4.

Mehrotra, P., Chatterjee, J., Chakraborty, C., Ghoshdastidar, B. and Ghoshdastidar, S. (2011). Automated screening of polycystic ovary syndrome using machine learning techniques, *2011 Annual IEEE India Conference* pp. 1–5.

Morang, M., Chasta, P. and Chandrul, M. (2019). A review on "polycystic overy syndrom pcos", *International Journal of Trend in Scientific Research and Development* **Volume-3**: 60–66.

Neuzil, A. (2014). What is polycycstic ovary syndrome (pcos)?
URL: *https://www.austinfitmagazine.com/November-2014/what-is-polycycstic-ovary-syndrome-pcos*

Purnama, B., Wisesti, U. N., Adiwijaya, Nhita, F., Gayatri, A. and Mutiah, T. (2015). A classification of polycystic ovary syndrome based on follicle detection of ultrasound images, *2015 3rd International Conference on Information and Communication Technology (ICoICT)* pp. 396–401.

Ramamoorthy, S., Kumar, T. S., Mansoorroomi, S. M. and Premnath, B. (2020). Enhancing intricate details of ultrasound pcod scan images using tailored anisotropic diffusion filter (tadf).

Rihana, S., Moussallem, H., Skaf, C. and Yaacoub, C. (2013). Automated algorithm for ovarian cysts detection in ultrasonogram, *2013 2nd International Conference on Advances in Biomedical Engineering* pp. 219–222.

Saito, H. and Ohmori, H. (2011). Control of an abnormal human menstrual cycle in pcos by speed gradient algorithm, *SICE Annual Conference 2011*, pp. 1436–1441.

Setiawati, E., Adiwijaya and Tjokorda, A. B. W. (2015). Particle swarm optimization on follicles segmentation to support pcos detection, *2015 3rd International Conference on Information and Communication Technology (ICoICT)* pp. 369–374.

Sitheswaran, R. and Malarkhodi, S. (2014). An effective automated system in follicle identification for polycystic ovary syndrome using ultrasound images, *2014 International Conference on Electronics and Communication Systems (ICECS)*, pp. 1–5.

Tan, Y., Zhang, C., Ma, Y. and Mao, Y. (2015). Knowledge discovery in databases based on deep neural networks, *2015 IEEE 10th Conference on Industrial Electronics and Applications (ICIEA)*, pp. 1217–1222.

TechBlogMU (2018). Describe the various functionalities of data mining as a step in the process of knowledge discovery.
URL: *https://techblogmu.blogspot.com/2018/05/describe-various-functionalities-of.html*

Vigil, P., Cortés Cortés, M., Zúñiga, A., Riquelme, J. and Ceric, F. (2009). Scanning electron and light microscopy study of the cervical mucus in women with polycystic ovary syndrome, *Journal of electron microscopy* **58**: 21–7.

Xie, N.-N., Wang, F.-F., Zhou, J., Liu, C. and Qu, F. (2020). Establishment and analysis of a combined diagnostic model of polycystic ovary syndrome with random forest and artificial neural network, *BioMed Research International* **2020**: 1–13.